



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 95244

TO: Manjunath N Rao
Location: cm1/10a11/10d01
Art Unit: 1652
Monday, June 09, 2003
Case Serial Number: 10/005499

From: Edward Hart
Location: Biotech-Chem Library
CM1-6B02
Phone: 305-9203

edward.hart@uspto.gov

Search Notes

Examiner Rao,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

THIS PAGE BLANK (USPTO)

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, .rnpm and .rnpn

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, .rapm and .rapn

The Pending database search results should not be left in the case because they contain data that is confidential.

THIS PAGE BLANK (USPTO)

95244

STIC-Biotech/ChemLib

Fr m: Rao, Manjunath N.
Sent: Friday, May 23, 2003 2:54 PM
T : STIC-Biotech/ChemLib
Subject: Sequence search request for 10/005,499

Fr m: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10D 01
Phone: 306-5681

Date: 5-23-03

Please search the following as soon as possible for application with serial number **10/005499**

1. SEQ ID NO: 377 against all commercial nucleic acid databases including issued patents database and pending application database and provide a print of all results.

2. SEQ ID NO:378 against all commercial protein databases including issued patents database and pending application database and provide a print of all results.

If y u have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
Bi technology Patent Examiner
Art Unit 1652, Room 10A11
Mail Box in 10D01
Crystal Mall 1, USPTO.

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 6/2/03
Date Completed: 6/9/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 01/02
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)



STIC SEARCH RESULTS

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library, CM1 – Circ. Desk



THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2003, 16:28:18 ; Search time 5224 Seconds
(without alignments)
11231.098 Million cell updates/sec

Title: US-10-005-499-377

Perfect score: 2016

Sequence: 1 aggaaggttagggagcgga.....tggggctacagctgtgctg 2016

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: en_ba:*

16: en_fun:*

17: en_hum:*

18: en_in:*

19: en_mu:*

20: en_om:*

21: en_or:*

22: en_ov:*

23: en_pat:*

24: en_ph:*

25: en_pl:*

26: en_ro:*

27: en_sts:*

28: en_un:*

29: en_vi:*

30: en_htg_hum:*

31: en_htg_inv:*

32: en_htg_other:*

33: en_htg_mus:*

34: en_htg_pln:*

35: en_htg_rtd:*

36: en_htg_mam:*

37: en_htg_vrt:*

38: en_sy:*

39: en_htgo_hum:*

40: en_htgo_mus:*

41: en_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1172.8	58.2	175186	2	AC026125 Homo sapi
2	1172.8	58.2	194578	2	AC125611 Homo sapi
3	1172.8	58.2	207607	2	AC010173 Homo sapi
4	661	32.8	177172	2	AC096892 Rattus no
5	661	32.8	179421	2	AC110690 Rattus no
6	623	30.9	221470	2	AC027679 Mus muscu
7	367.6	18.2	1526	9	BC008798 Homo sapi
8	358.6	17.8	1297	9	AF095154 Homo sapi
9	328.2	16.3	777	10	AF095155 Mus muscu
10	305.8	15.2	2526	10	AB044560 Mus muscu
11	296.8	14.7	970	10	BC024634 Mus muscu
12	278	13.8	847	10	BC022724 Mus muscu
13	271.6	13.5	1742	6	AX463516 Sequence
14	259.2	12.9	981	6	AX456049 Sequence
15	258.8	12.8	795	6	AX477323 Sequence
16	237.8	11.8	750	6	AX477325 Sequence
17	222	11.0	52065	9	AL353576 Human DNA
18	220.4	10.9	195191	9	AC024339 Homo sapi
19	207.8	10.3	198046	2	AL731670 Mus muscu
20	205.4	10.2	10184	9	AF410771 Homo sapi
21	205.4	10.2	204275	2	AC015936 Homo sapi
22	203.4	10.1	11031	10	AB045983 Mus muscu
23	157.2	7.8	81619	5	AL672065 Zebrafish
24	152.2	7.5	185332	9	AC016673 Homo sapi
25	151.4	7.5	94925	2	AC127906 Rattus no
26	146.6	7.3	191804	2	AC019017 Homo sapi
27	139.8	6.9	674	9	HS342389
28	133	6.6	190487	2	AC084310 Mus muscu
29	131.4	6.5	94925	2	AC127906 Rattus no
30	105.6	5.2	162599	9	AL360230 Human DNA
31	100	5.0	1056	10	MUSBCLQ
32	100	5.0	1059	6	AX305847
33	100	5.0	1059	6	MUSC1Q
34	98.2	4.9	1136	6	AX401995
35	98.2	4.9	1136	10	RNC1QB
36	95.8	4.8	106037	2	AL158086 Homo sapi
37	95.8	4.8	171553	2	AC013611 Homo sapi
38	95.8	4.8	176558	2	AL611946 Homo sapi
39	95.8	4.8	185848	2	AC025929 Homo sapi
40	95.6	4.7	173239	2	AC113790 Rattus no
41	94.2	4.7	977	9	HUMC1QB2
42	93.8	4.7	182829	10	AL627214 Mouse DNA
43	91	4.5	935	9	HSC1QBR
44	91	4.5	1076	9	BC008983
45	90.4	4.5	1357	6	AX068378 Sequence

ALIGNMENTS

RESULT 1
AC026125
LOCUS
DEFINITION
AC026125
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AC026125
Homo sapiens chromosome 12 clone RP11-234P5, WORKING DRAFT
SEQUENCE, 14 unordered pieces.
AC026125.22 GI:21430995
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 175186)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

AC026125
175186 bp DNA linear HTG 18-JUN-2002
Homo sapiens chromosome 12 clone RP11-234P5, WORKING DRAFT
SEQUENCE, 14 unordered pieces.
AC026125.22 GI:21430995
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 175186)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 194578: contig of 194578 bp in length.
FEATURES
source
Location/Qualifiers
1..194578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-161H23"
BASE COUNT 54725 a 45940 c 44426 g 49485 t 2 others
ORIGIN
Query Match 58.2%; Score 1172.8; DB 2; Length 194578;
Best Local Similarity 99.5%; Pred. No. 3e-157;
Matches 1218; Conservative 0; Mismatches 2; Indels 4; Gaps 4;
QY 1 AGAAGGTTAGGAGCGGAGAGGACCGGCCCGCA-GAGAGAGCGCGGGCCAGGGCT 59
Db 110407 AGAAGGTTAGGAGCGGAGAGGACCGGCCCGCAAGGAGAGAGGCGGCGGCGGCT 110466
QY 60 ACTAGCAGGACTGGGGCGCGCAGGGGTAGCAAGGTGAGTGGTGTGCTTGCACAGGC 119
Db 110467 -TAGCAGGAGTGGGGCGCGCAGGGGTAGCAAGGTGAGTGGTGTGCTTGCACAGGC 110525
QY 120 AGAGCGCAAACTACTAGGAGATCGCGCCCGGTAGCAGCAGCCCGCAGCTCAGAGCCCG 179
Db 110526 AGAGCGCAAACTACTAGGAGATCGCGCCCGGTAGCAGCAGCCCGCAGCTCAGAGCCCG 110585
QY 180 GGAGTTCGGAGCGCGGAGCAGTCCCTCTCCATCAGGAGTGTGTATCTGGGCAGT 239
Db 110586 GGAGTTCGGAGCGCGGAGCAGTCCCTCTCCATCAGGAGTGTGTATCTGGGCAGT 110645
QY 240 CTGGGACCCAGGACCGCGCATCCCTGAGAGAGCAGCAGTCTGGAGAGCAGGCATCTCA 299
Db 110646 CTGGGACCCAGGACCGCGCATCCCTGAGAGAGCAGCAGTCTGGAGAGCAGGCATCTCA 110705
QY 300 GATCCCTAAGAAACCCGCTCCGAGAACCGCGGATCTCAGGTGCCCGCAGGATCGTTAGG 359
Db 110706 GATCCCTAAGAAACCCGCTCCGAGAACCGCGGATCTCAGGTGCCCGCAGGATCGTTAGG 110765
QY 360 ACTGAACGGAGGGTACTAGAGGACCACTGGCTCTCGACCGTGGGAGCTGCCCTGACG 419
Db 110766 ACTGAACGGAGGGTACTAGAGGACCACTGGCTCTCGACCGTGGGAGCTGCCCTGACG 110825
QY 420 TAACCCACAGGGGCTCCCTTTGACGAGCGCTTGGGAGCGGCACCGCG-GCCTGGA 478
Db 110826 TAACCCACAGGGGCTCCCTTTGACGAGCGCTTGGGAGCGGCACCGCGCGGCTGGA 110885
QY 479 GCCCGCAGAGGAGGTAAGGGAGCGGGGCGCAGCGCTCGGGGGAGTGCAGACCCAGGC 538
Db 110886 GCCCGCAGAGGAGGTAAGGGAGCGGGGCGCAGCGCTCGGGGGAGTGCAGACCCAGGC 110945
QY 539 CCAAGCGGGTACCG-CTCCTGGCCCGGGAGAGCCCGCGCCCGCGCAGCCATTGCGGC 597
Db 110946 CCAAGCGGGTACCGCCCTCTGGCCCGGGAGAGCCCGCGCCCGCGCAGCCATTGCGGC 111005
QY 598 CRAAGCTGAGGAAGATTGCTGCCCTGCGCAGGCTCGCGCTCAGCGCGGCGCAAGAGGT 657
Db 111006 CRAAGCTGAGGAAGATTGCTGCCCTGCGCAGGCTCGCGCTCAGCGCGGCGCAAGAGGT 111065
QY 658 GCGGGCGGGCGGCTCGGAGTGCCCATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717
Db 111066 GCGGGCGGGCGGCTCGGAGTGCCCATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 111125

QY 718 GGTGCACAGCTCCCGGGGCGGCGCAGCTACGAGATGCTGGGTGCTGCGGCATGCTGTG 777
Db 111126 GGTGCACAGCTCCCGGGGCGGCGCAGCTACGAGATGCTGGGTGCTGCGGCATGCTGTG 111185
QY 778 CGACCCGCGATGGCCCGCGTGGCCCTGCCGAGCGGCGGCGCTTCGTCGTCGCCCTT 837
Db 111186 CGACCCGCGATGGCCCGCGTGGCCCTGGTCCCGACGCGGCGGCTTCGTCGTCGCCCTT 111245
QY 838 CCGCCGAGCGCCCAAGGAGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 897
Db 111246 CCGCCGAGCGCCCAAGGAGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 111305
QY 898 TGGACCACAGGTTCGAAGAGGCGCCCGCAGAGAGAACCGCGGCGGCGGCGGCGGCGGCGG 957
Db 111306 TGGACCACAGGTTCGAAGAGGCGCCCGCAGAGAGAACCGCGGCGGCGGCGGCGGCGGCGG 111365
QY 958 TCCCGTCCAGGTCCCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1017
Db 111366 TCCCGTCCAGGTCCCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 111425
QY 1018 CTACGCGGCGCTGCGGCGGCGCCACAGAGGTTACGAGGTGCTGCGGCTTCGACGACGTGT 1077
Db 111426 CTACGCGGCGCTGCGGCGGCGCCACAGAGGTTACGAGGTGCTGCGGCTTCGACGACGTGT 111485
QY 1078 GACCAACGTGGGCAAGCCCTACGAGGCGGCGGCGGCAAGTTACTTGCCCATGCGCAGG 1137
Db 111486 GACCAACGTGGGCAAGCCCTACGAGGCGGCGGCGGCAAGTTACTTGCCCATGCGCAGG 111545
QY 1138 CGTCTACTTCTTCCGTTTACACGTGCTATCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1197
Db 111546 CGTCTACTTCTTCCGTTTACACGTGCTATCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 111605
QY 1198 CGACCTCATGAAGACGGACAGGT 1221
Db 111606 CGACCTCATGAAGACGGACAGGT 111629
RESULT 3
AC010173/c 207607 bp DNA linear HTG 31-JUL-2002
LOCUS Homo sapiens chromosome 12 clone RP11-977B10, WORKING DRAFT
DEFINITION SEQUENCE, 8 unordered pieces.
AC010173 AC010173 GI:220032320
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 207607)
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Anaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbiera, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhey, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Devilla, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Fallis, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A., Hernandez, J.,
Hernandez, O., Hodgeson, A., Hogue, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W., Loulsegue, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,


```
OY 838 CCGCCAGGCGCCAAAGGAGAGGTGGCGCGCGCGGAAAGAGGAGGCTCGCGGGGCCCC 897
Db 49040 CCGCCAGGCGCCAAAGGAGAGGTGGCGCGCGCGGAAAGAGGAGGCTCGCGGGGCCCC 48981
OY 898 TGGACACACAGGTCCAAAGAGGCCCCCAGGAGAACCCGCGAGCCAGGCCCCCGGGGCCCC 957
Db 48980 TGGACACACAGGTCCAAAGAGGCCCCCAGGAGAACCCGCGAGCCAGGCCCCCGGGGCCCC 48921
OY 958 TCCCGGTTCAGGTTCGCGGGGGGTGGCGCGCGCTGCGCGGTACGTCGCTCGCATTCCTTT 1017
Db 48920 TCCCGGTTCAGGTTCGCGGGGGGTGGCGCGCGCTGCGCGGTACGTCGCTCGCATTCCTTT 48861
OY 1018 CTACGCGGCGCTCGCGCGCCCCACGAGGTTTACGAGGTCTCGCGCTTCGACGACGTGT 1077
Db 48860 CTACGCGGCGCTCGCGCGCCCCACGAGGTTTACGAGGTCTCGCGCTTCGACGACGTGT 48801
OY 1078 GACCAACGTGGCAACGCTACGAGCGCAGCCAGGCAAGTTTACTTGGCCCAATGCCAGG 1137
Db 48800 GACCAACGTGGCAACGCTACGAGCGCAGCCAGGCAAGTTTACTTGGCCCAATGCCAGG 48741
OY 1138 CGTCTACTTCTTCCGCTTACACGTCTCATGCGCGCGCGGCGGACGACGATGTGGGC 1197
Db 48740 CGTCTACTTCTTCCGCTTACACGTCTCATGCGCGCGCGGCGGACGACGATGTGGGC 48681
OY 1198 CGACCTCATTGAGGAAGGACGACAGT 1221
Db 48680 CGACCTCATTGAGGAAGGACGACAGT 48657
```

RESULT 4

AC096892/c

LOCUS

DEFINITION

Rattus norvegicus clone CH230-68F12, *** SEQUENCING IN PROGRESS

***, 74 unordered pieces.

AC096892

AC096892.5 GI:21728994

HTG; HTGS_PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 17172)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Bowie,L., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,

Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,

Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,

Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,

Kratochvil,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H.,

Lozadó,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

Massey,E., Mawhinney,E., McLeod,M.P., Meador,W., Mei,G., Metzker,M.,

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,

Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,

Nguyen,N., Nickerson,E., Nwokendo,S., Ogih,M., Okunolu,G.,

Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,

Rattus.

1 (bases 1 to 17172)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Bowie,L., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,X.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
2 (bases 1 to 17172)
Worley,K.C.
Direct Submission
Submitted (30-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 17172)
Worley,K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17972984.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGWU
Center clone name: CH230-68F12
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 114650 bases at least Q40
Consensus quality: 121322 bases at least Q30
Consensus quality: 126307 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 74 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1047: contig of 1047 bp in length
* 1048 1147: gap of unknown length
* 1148 2776: contig of 1629 bp in length
* 2777 2876: gap of unknown length
* 2877 4334: contig of 1458 bp in length
* 4335 4434: gap of unknown length
* 4435 5853: contig of 1419 bp in length
* 5854 5953: gap of unknown length
* 5954 7150: contig of 1197 bp in length
* 7151 7250: gap of unknown length
* 7251 8632: contig of 1382 bp in length
* 8633 8732: gap of unknown length
* 8733 10281: contig of 1549 bp in length
* 10282 10381: gap of unknown length
* 10382 12045: contig of 1664 bp in length
* 12046 12145: gap of unknown length
* 12146 13156: contig of 1011 bp in length
* 13157 13256: gap of unknown length
* 13257 14271: contig of 1014 bp in length
* 14271 14370: gap of unknown length
* 14371 15797: contig of 1427 bp in length
* 15798 15897: gap of unknown length

15898 17311: contig of 1414 bp in length
17411: gap of unknown length
17412 17411: gap of unknown length
17412 18905: contig of 1494 bp in length
18906 19005: gap of unknown length
19006 20382: contig of 1377 bp in length
20382 20482: gap of unknown length
20482 22778: contig of 1796 bp in length
22778 22378: gap of unknown length
22378 23632: contig of 1254 bp in length
23632 23732: gap of unknown length
23732 25520: contig of 1788 bp in length
25520 25620: gap of unknown length
25620 26282: contig of 1308 bp in length
26282 27028: gap of unknown length
27028 28783: contig of 1755 bp in length
28783 28884: gap of unknown length
28884 30947: contig of 2064 bp in length
30947 31047: gap of unknown length
31047 32504: contig of 1457 bp in length
32504 32604: gap of unknown length
32604 34035: contig of 1431 bp in length
34035 34135: gap of unknown length
34135 36108: contig of 1873 bp in length
36108 37781: gap of unknown length
37781 39370: contig of 1489 bp in length
39370 39470: gap of unknown length
39470 40743: contig of 1273 bp in length
40743 40843: gap of unknown length
40843 42540: contig of 1697 bp in length
42540 42640: gap of unknown length
42640 43652: contig of 1012 bp in length
43652 43752: gap of unknown length
43752 46346: contig of 2594 bp in length
46346 46446: gap of unknown length
46446 48238: contig of 1792 bp in length
48238 48338: gap of unknown length
48338 49764: contig of 1426 bp in length
49764 49864: gap of unknown length
49864 51020: contig of 1156 bp in length
51020 51120: gap of unknown length
51120 52762: contig of 1642 bp in length
52762 52862: gap of unknown length
52862 54944: contig of 2082 bp in length
54944 55044: gap of unknown length
55044 56511: contig of 1467 bp in length
56511 56811: gap of unknown length
56811 58449: contig of 1838 bp in length
58449 58549: gap of unknown length
58549 60540: contig of 1991 bp in length
60540 60640: gap of unknown length
60640 62264: contig of 1624 bp in length
62264 62364: gap of unknown length
62364 64155: contig of 1791 bp in length
64155 64255: gap of unknown length
64255 66511: contig of 2256 bp in length
66511 66611: gap of unknown length
66611 67983: contig of 1372 bp in length
67983 68083: gap of unknown length
68083 70381: gap of unknown length
70381 73125: contig of 2744 bp in length
73125 73225: gap of unknown length
73225 75265: contig of 2040 bp in length
75265 75365: gap of unknown length
75365 78038: contig of 2673 bp in length
78038 78138: gap of unknown length
78138 80049: contig of 1911 bp in length
80049 80149: gap of unknown length
80149 82178: contig of 2029 bp in length
82178 82279: gap of unknown length
82279 85330: contig of 3052 bp in length

* 85331 85430: gap of unknown length
* 85431 87933: contig of 2503 bp in length
* 87934 88033: gap of unknown length
* 88034 89126: contig of 1093 bp in length
* 89127 89226: gap of unknown length
* 89227 91784: contig of 2558 bp in length
* 91785 91884: gap of unknown length
* 91885 94448: contig of 2564 bp in length
* 94449 94548: gap of unknown length
* 94549 97418: contig of 2870 bp in length
* 97419 97518: gap of unknown length
* 97519 101135: contig of 3617 bp in length

Query Match 32.8%; Score 661; DB 2; Length 177172;
Best Local Similarity 75.7%; Pred. No. 8.2e-85;
Matches 936; Conservative 0; Mismatches 265; Indels 36; Gaps 8;

QY 1 AGGAAGTTAGGAGCGGAGAGGACCGCGCCGCG-AGAGAGAGCGCGGCGGCGAGGCT 59
DB 65894 AGGAAGTTAGGAGCGGAGAGGAGCGCTGCGCGGAGAGGAGCGCTGGCTAGGCG 65835
QY 60 ACTAGCAGGAGCTGGGCGCGCGAGGAGGCTAGCAAGGTGAGTGGTTCGCAAGAGCG 119
DB 65834 TCTGAGGAGGCTGCGCGCCCGCTGCCGAAGAGGCG--AGGTAGTGTCTTCCACAGCT 65777
QY 120 AGAGCGCAAAACCTACTAGGAGATCGCGCCGCTGAG-----CAGCACCGCGAGCTC 171
DB 65776 AGAGTGCRAAATCTCGAGAGGTCTCATCCAGTGAGAGTCCCCCAACCGCGCAGCTC 65717
QY 172 AGAGCGCGGAGCTCGGAGCGCGGAGAGTCCCTCTTCCATCAGGAGAGTGGTGTATC 231
DB 65716 AGAGCTGGGAGCTCGGAGAGGAGAGGAGTCCCTGCTAGCGCGGAGAAATCAAGC 65657
QY 232 TGGCGAGTCTGGACCCAGGCGACCGCCATCCCTGAGAGAGCAGCAGTCTGGAGAGCAG 291
DB 65656 AGGGGTGTCGGGAGCCAGACAC--TGCCATCCGAAGAAGAACACTGTCTGGAGAGTG 65599
QY 292 GCATCTCAGATCCCTAAGAAACAGCGCTCCGAGAGCGCGGATCTCAGGTGCCCGAGGA 351
DB 65598 GCATCTTAGACCTGTGAGAAACAGCGCTCCGAAGAGTGTGAGGCTCAGGCGCCCTGGA 65539
QY 352 TCGTTAGGACTGAACGGGAGGAGTACTAGAGAGCAGTGTGCTCTGGACCTCGGGAGCTGC 411
DB 65538 TCATTAGAGCTGCACCGCGAGGAGTAATATCAGACTTAAGTCTTGAACCTCAAGGCTGT 65479
QY 412 CCTGACGTAACCCACGAGG--GCCTCCCTTACGACAGCGCTTGGGAGCGGACCG 468
DB 65478 CCTTGGCCCACTCAGCGGGGAGATCGCCACCTGACAGCGGTTTGGGGAACAGCACTG 65419
QY 469 CCGGCTTGAGCCCGCAGAGGAGGTAAGGGAGCGGGGAGCGCGGCGAGCGCTCGGGGAGTGC 528
DB 65418 CTGTC--ACAGGAGGAAATTCACCGCGGCAAGGTGGTTCGGGACACGCGTCGGGAGA-TCC 65362
QY 529 AGACCCAGGCGCAAGCGGAGTACCGCTCTTGGCGCGGAGAGCGCGCGCGCGCGCAGC 588
DB 65361 AGACCGGGCGCGAGCGCGCGGAGAGC-----CCCGGCGCTCCCGCGC 65319
QY 589 CATTCGGCCCAAGAGTGAAGAGATTTGCTGGCCCTTGGCAGCGTGCAGCGCTGAGCGCGG 648
DB 65318 CATTCGGCCCAAGTGAAGAGATTTGCTGGCCCTTGGCAGCGTGCAGCGCTGAGCGCGG 65259
QY 649 CAAGAGGTTGGCGCGCGCGCTCGGAGTGGCCATGTTGCTGCTGCTGCTGCTGCTGCTGCT 708
DB 65258 CAACCGGGTGGCGCGCGAGCGCTCGGGTGGCCATGTTGCTGCTGCTGCTGCTGCTGCTGCT 65199
QY 709 CCCGCTGCTGTCACAGCTCCCGGGGCGGAGCGGAGTACAGAGATGCTGGGTCGTCGCGC 768
DB 65198 CCCGCTACTGTCACAGCTCTCGGGGACCAACACTATGAGATGCTAGTGGCTGTGCG 65139
QY 769 CATGTTGTCGACCGCGATGGCGCGCGCTGGTCCCGAGCGCGCGCTGCTGCTGCTGCTGCT 828
DB 65138 CATGTTGTCGACCGCGATGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 65079

```
QY 829 GCCCCCTTCCCGCCAGGCGCCAAAGGAGAGGTGGCGCGCGGCGGAAAGACAGGCGCTGCG 888
D 829 GCCCCCTTCCCGCCAGGCGCCAAAGGAGAGGTGGCGCGCGGCGGAAAGACAGGCGCTGCG 888
Db 65078 GCCTCCCTTCCCTCCAGGTGCGCCAAAGGAGAGGTAGGCGCGGCGGAAAGCGGCGCTGCG 65019
QY 889 GGGGGCCCCCTGGACACCAAGGTCCAGAGGGCCCCCAGGAGACCGCGGAGCCAGGCC 948
D 889 GGGGGCCCCCTGGACACCAAGGTCCAGAGGGCCCCCAGGAGACCGCGGAGCCAGGCCAG 948
Db 65018 GGGACCCCGGAGGAGCCCGGGTCCAGAGGGCCCCCAGGAGACCGCGGAGCCAGGCCAG 64959
QY 949 CCGGGGCGCTCCCGGGTCCAGGTCCGGGCGGGGCGCGCCCGCTGCGCGGTACGTCGCTCG 1008
D 949 CCGGGGCGCTCCCGGGTCCAGGTCCGGGCGGGGCGCGCCCGCTGCGCGGTACGTCGCTCG 1008
Db 64958 TCAGGTCCCTGCGGCGCCAGGCGCTGGAGGGGCGCGCCCGCTGCGCGGTATGATCCCGG 64899
QY 1009 CATTGCTTCTTACGCGGGCGCTGGCGGCGCCCGACGAGGTTAGAGGTGCTGCGCTTCGA 1068
D 1009 CATTGCTTCTTACGCGGGCGCTGGCGGCGCCCGACGAGGTTAGAGGTGCTGCGCTTCGA 1068
Db 64898 AATTGCTTCTTATGCGGGCTAAGCGGCTCATGAGGGTTATGAGGTGCTGCGCTTGA 64839
QY 1069 CGAGGTGGTACCAAGGTGGGCAACGCTTACGAGGCGAGCGGCGGCAAGTTACTTGGCC 1128
D 1069 CGAGGTGGTACCAAGGTGGGCAACGCTTACGAGGCGAGCGGCGGCAAGTTACTTGGCC 1128
Db 64838 CGAGGTGGTACCAAGGTGGGCAACGCTTACGAGGCGAGCGGCGGCAAGTTACTTGGCC 64779
QY 1129 CATGCCAGGCGTCTTCTTCTTCCCTTACCAGTGTCTCATGCGCGGCGGCGGACCGACCG 1188
D 1129 CATGCCAGGCGTCTTCTTCTTCCCTTACCAGTGTCTCATGCGCGGCGGCGGACCGACCG 1188
Db 64778 CATGCCAGGCGTCTTCTTCTTCCCTTACCAGTGTCTCATGCGCGGCGGCGGACCGACCG 64719
QY 1189 CATGTGGCGGACCTCATGAAGAACGAGACGAGTCCGG 1225
D 1189 CATGTGGCGGACCTCATGAAGAACGAGACGAGTCCGG 1225
Db 64718 CATGTGGCGGACCTCATGAAGAACGAGACGAGTCCGG 64682
```

RESULT 5

```
AC110690
LOCUS Rattus norvegicus clone CH230-290M8, 179421 bp DNA linear HTG 13-JUL-2002
DEFINITION *** 29 unordered pieces.
AC110690
VERSION AC110690.4 GI:21740747
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 179421)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J.J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homsf., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Liep,C., Liu,J., Liu,W., Lousegh,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenko,S., Ogih,M., Okuwon,G.,
Oragunye,N., Oyedoro,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Picken,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
```

```
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaik,E., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 179421)
Worley,K.C.
Direct Submission
Submitted (15-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 179421)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:20514409.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRM
Center clone name: CH230-290M8
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 147907 bases at least Q40
Consensus quality: 151353 bases at least Q30
Consensus quality: 153945 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1305: contig of 1305 bp in length
* 1306: gap of unknown length
* 1406: contig of 1680 bp in length
* 3085: contig of 1680 bp in length
* 3185: gap of unknown length
* 4216: contig of 1031 bp in length
* 4316: gap of unknown length
* 4217: contig of 1579 bp in length
* 5895: gap of unknown length
* 5996: contig of 1011 bp in length
* 7007: gap of unknown length
* 7106: contig of 1700 bp in length
* 8906: gap of unknown length
* 10034: contig of 1128 bp in length
* 10134: gap of unknown length
* 11868: contig of 1734 bp in length
* 11969: gap of unknown length
* 13610: contig of 1642 bp in length
* 13710: gap of unknown length
* 13711: contig of 1375 bp in length
* 15085: gap of unknown length
* 15185: gap of unknown length
* 16830: contig of 1645 bp in length
* 16931: gap of unknown length
```


Strausberg,R. Direct Submission Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgabbs-re@mail.nih.gov Tissue procurement: ATCC CDNA Library Preparation: Rubin Laboratory DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@hgri.nih.gov Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G., Brinkley,C., Brooks,S., Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R., Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J., Tionson,E.E., Touchman,J.W., Tsugeon,C., Vogt,J.L., Walker,M.A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAL Plate: 12 Row: f Column: 8 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 3747096.	Location/Qualifiers 1. .l526 /organism="Homo sapiens" /db_xref="LocustID:l0882" /db_xref="Taxon:9606" /clone="MGC:3776 IMAGE:3635430" /tissue_type="Placenta, choriocarcinoma" /clone_lib="NIH_MGC_21" /lab_host="DH10B-R" /note="Vector: pOTB7" 238. .l014 /codon_start=1 /product="Clq-related factor" protein_id="AAH08798.1" /translation="MLLVVLPVLVSSGGPEGHYEMLTGTCMVDPYPARGPGAGA RTDGDALSESGAPPPSTLVQGGKRGPKGTGKPGPDGPDPGVPVGPKEKGEF GKPGPPGLPAGAGGAISTATITTPRVAFYALKNPHEGYEVLKFDVVNTLNNGNDY RASRGTETICGTFTFYTHVMRGDDGTSMWADLKNGOVRSASIAODADQNTDYASN SVILHDAGDFVIKLDGKAHGNSKNKSTFSFIYS"
AUTHORS TITLE JOURNAL	REMARK COMMENT
Db 154229 CCCTGAATCATTTAGAACCTGCACCCAGAGGTATAATCAGACTGTAGGTTCTGAACCCCTCAA 154288	Qy 405 GAGCTGCCCTCACGTAACCCACGAGGG--CCTCCCCTTGACGAGCGGTGCGGGAGCG 462
Db 154289 GGCTGTCTTGGCCCCAGCTACGGGGAGATGCCACCCTGACAGGGCGTTGGGGAACA 154348	Qy 463 GCACCGCCGCTGGAGCCCGCAGAGGCAAGGTAAAGGGAGCGGGGGGACGCCCTCGGG 522
Db 154349 GCACCTGCTGC--ACAGAAGAAATTCGCCGGCAAGGTGTCGGGCACAGCCGTCGGGA 154406	Qy 523 GAGTGCAGACCCAGCCCAAGCGGGTACCGCTCTGCTGCGCCGGGAGAGCCCGGGCCC 582
Db 154407 GA-TCCAGACCGGGCCGAGCCGG-----GAGAGCCCGGGGGCCC 154447	Qy 583 GGCAGCATTTGCGCCCAAGAGTAGAAGATTGCTGCTGCGCTGCGAGCTGCGGCTGAG 642
Db 154448 GCCACCATTTGCGCCCAAGTAGAGAGATTGCTGCGCTGCGCGGTGACAGGCTGAG 154507	Qy 643 CGGGCGCAAGAGGGTGGGGGGCGCGCGCTGCGAGTGGCCATGCTGCTGCTGCTGGT 702
Db 154508 -GCCCGCAAGCAGTGGCGGGCGCACCTCGAGGGTGGCCATGCTGCTGCTGCTGGT 154566	Qy 703 GGCATCCGCTGCTGGTGCACAGCTCCCGGGGCGCGCACGACATACAGATGCTGGGTG 762
Db 154567 AGCCATTCGCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 154626	Qy 763 CTGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 822
Db 154627 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 154686	Qy 823 TTCCGTCGCCCCCTTCCGCGCCAGGCGCGCAAGGAGAGGTGGCGCGCGGAAAGCAGG 882
Db 154687 CTCGGTGCCTTCTTACCTCCAGGTGCCAAGGAGAGGTGGCGCGGAGGAGGAGG 154746	Qy 883 CCTGGGGGGCCCCCTGACACACAGTCTCAAGAGGGCCCCCAGAGAGAACCCGCGACGCC 942
Db 154747 CTTGGGGGGGGCCACCAAGAGACCCAGCTCCAGAGGGCGCCCTGGAGAGCGAGCG 154806	Qy 943 AGSCCCCGCGCGCTTCCCGGTCACAGTCCGGGGGGGTGGCGCGCGCTGCGGCTACGT 1002
Db 154807 AGGACCCCAAGTCTCTCTGCGCCAGGCCCTGGAGGGCAGCGCCCTGCGCGCTATG 154866	Qy 1003 GCTCGCATGCTTTCTAOGCGGGGCTCGCGCGGCCCAAGAGGTTACGAGTGTCTGCG 1062
Db 154867 ACCCGAATTCCTTCTAGCGGGTCTAAGCGGGCTTACGAGGGTTATGAGTGTGCG 154926	Qy 1063 CTTGACGACGTGTTGACCAAGTGGGCAAGCTTACGAGCGAGCGAGCGCAAGTTTAC 1122
Db 154927 CTTTACGACGTGTTGACCAAGTGGGCAAGCTTACGAGCGAGCGAGCGCAAGTTTAC 154986	Qy 1123 TTGCCCCATGCCAGCGTCTACTCTTCTGCTTACACGTGCTCATGCGCGCGCGACGG 1182
Db 154987 GTGCCCCATGCCAGAGTCTATTTCTTCTGCTTACATGCTCATGCGAGCGGTGACGG 155046	Qy 1183 CACCAAGCATGTGGCGGCGCTCATGAAGAGCGAGAGTTCGG 1225
Db 155047 TACCAGCATGTGGCGGATCTGATGAAGAGCGAGGTGAGG 155089	Db 155047 TACCAGCATGTGGCGGATCTGATGAAGAGCGAGGTGAGG 155089
RESULT 7	BC008798
LOCUS	BC008798.1 GI:14250665
DEFINITION	Homo sapiens, Clq-related factor, clone MGC:3776 IMAGE:3635430, mRNA, complete cds.
ACCESSION	BC008798
VERSION	BC008798.1
KEYWORDS	MGC.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE 1 (bases 1 to 1526)	

Qy	849	CC	AAGGAGAGGTGGCGCGCGGGAAGACAGAGCTGCGGGGGCCCCCTTGGACCAACCAG	908
Db	440	CC	AGGGGAAGCGGGCGCGCACCGGAAGCCGGGCGCTCCGGGGCGCTCCCGGGAGACCCAG	499
Qy	909	GT	CAAGAGGGCC-----CCACGAGAACCCGACGAGCCAGGCCGCC	950
Db	500	GT	CTCCCGGCCCTGTGGGCGCCCGGGGGAGAGGGTGAGCCAGGACCGGGGCCCTC	559
Qy	951	CG	GGGCCCTCCCGGTCACAGGTCCGGGGGGGGTGGCGCCGCTGCGGGCTAC-----GTGC	1004
Db	560	CG	GGGGTTCGCGGGCGGGGCGAGCGGCCATCAGCACTGCCACCTACACACACGGTGC	619
Qy	1005	CT	CGCATTTGCTTACGCGGGGCTCGGGGGGCCCCACGAGGGTTACGAGGTGCTCGGCT	1064
Db	620	CG	CGGTGGGCTTCTACGCGGGGCTCAAGAAACCCCCACGAGGGTTACGAGGTACTCAAGT	679
Qy	1065	TC	GACGACGTGGTGACCAAGCTGGGGAAACGCTTACGAGGCAGCGACGCGCAAGTTTACTTT	1124
Db	680	TT	GACGAGGTGTACCAACCTAGGCAACAACCTACGACGCGGCCAGGCGCAAGTTTACGT	739
Qy	1125	GC	CCCATGCCAGGGTCTACTTCTTGCCTTACACAGTGTCTATGCGGGGGGGGACGGCA	1184
Db	740	GC	AACATTCGCGGCACTACTTTTACCTTACCATGTCTCTATGCTGCGGGGGACGGCA	799
Qy	1185	CC	AGCATGTGGGCGGACCTCATGAAGAACGACAGTTCGGGGCACGCGCATTTGCTCAGG	1244
Db	800	CC	AGTATGTGGGCAGACCTCTGCAAGATGGCCAGGTGCGGGCCAGTGTATTGCCCAGG	859
Qy	1245	AC	CGGACCAAGAACTAGGACTACGCCAGCAACAGCGTCTATTGCACTTGGAGCGTGGCG	1304
Db	860	AC	CGGACCAAGAACTAGGACTACGCCAGCAACAGCGTGTCTCTGCACTTGGACGCGCGG	919
Qy	1305	AC	GAGTCTTTCATCAAGCTGAGCGGGGAAAGTGACAGCGCGCAACACCAACAAGTACA	1364
Db	920	AC	GAGTCTTTCATCAAGCTGAGTGGAGGCAAAACACAGCGGCGCAACAGCAAAATACA	979
Qy	1365	GC	ACTTCTCCGGCTTCATCATCTACCCGACTGAGCGGGCCCCCGCTGCCCCCGCT	1424
Db	980	GC	AGTCTCTGCGTTCATCATCTACTCCGACTGAGTCCCCACAGTCTCCCTCCACCCAC	1039
Qy	1425	CG	CCCCCTTCCTCCGTCCTACCACTCTCGCGCGCCCCACCGGAGGCGCACACCCAC	1484
Db	1040	GT	CCCTCACCGCGGGGTCCCTTCGGGGGGGCGACGATGACTGCGCCCTCGGCCAC	1099
Qy	1485	CC	TTTGAGAGCCTGGC	1500
Db	1100	CG	CTCGCTGCGCGGCG	1115

RESULT 8

AF095154
LOCUS AF095154 1297 bp mRNA linear PRI 15-OCT-1998
DEFINITION Homo sapiens C1q-related factor mRNA, complete cds.
ACCESSION AF095154
VERSION AF095154.1 GI:3747096

KEYWORDS
SOURCE
Homo sapiens.

ORGANISM Homo sapiens

• Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL

BOOKED
REFERENCE

AUTHORS Berube, N.G., Swanson, H.X. and Pereira-Smith, O.M.

TITLE
Direct Submission
Submitted (28-SEP-1998) Huffington Center on Aging, Baylor College
of Medicine, One Baylor Plaza, Houston, TX 77030, USA

FEATURES

source	1. .1297
--------	----------

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q21"
14. .790
/notes="CRF: contains a triple helical collagenous domain
and a Clq signature domain; highly expressed in the
brainstem"
/codon_start=1
/product="Clq-related factor"
/protein_id="AAG64186.1"
/db_xref="GI:3747097"
/translation="MLLVLVLPVLVSSGPGSGHVEMLGTCRMVCDPYAPARGPAGGA
RTGGDALSQSGAGPPSTLVQSPQSKRGKPGPGDGPDPGPGVPVGPQGEKGP
KGKPGPLPAGGSGALSTATYTVFVPAFYAGLKNKPHEGIEVLKDFDDVVYINLGNND
AASGKFTCNIPGTFYFTYVLMRGDGTSMADLCKNGQVRASAIQADADQNDYASN
SVTLHLDAQDEVIKLDGGRAKGGNGNKSTFTSGFIYSD"
218 a 484 c 406 g 189 t
BASE COUNT
ORIGIN

```

Query Match	17.8%;	Score 358.6;	DB 9;	Length 1297;
Best Local Similarity	66.4%;	Pred. No. 1.3e-41;		
Matches 592;	Conservative 0;	Mismatches 239;	Indels 60;	Gaps 3;

Qy	670	CGTCGGAGTGGCATTGCTGCTGCTGCTGCTGGCCATCCGCTGCTGCTGTCACAGCTC	729
Db	1		60
Qy	730	CCGCGCGCCAGCGCACTACGAGATGCTGGGTGCTGCCGATGCTGTGCGACCCGATGG	789
Db	61		120
Qy	790	GCCCCGTGGCCCTGGTCCGACG-----G	813
Db	121		180
Qy	814	CGGCCCTGCTTCGTTGCTGCCCTTCCCGCCAGCGCCAAAGGGAGAGTGGCGCGCGG	873
Db	181		240
Qy	874	GAAGAAGGCTTGGGGGGCCCCCTTGACACCAAGTCTCAAGAGGGGCCCA-----	925
Db	241		300
Qy	926	-----GGAGAACCCGCGACGGCCAGCGCCCCCGGGCCCTCCGCGCCCTGTGGGCGCGC	975
Db	301		360
Qy	976	CGGGGTGGCGCCGCTCCCGCTAC-----GTGCTTCGCATTGCTTTCACGCGGCGCT	1029
Db	361		420
Qy	1030	CGCGCGCCCCACGAGGGTTACGAGTGTCTGCGCTTCGACGAGCTGTGTGACCAACGTGGG	1089
Db	421		480
Qy	1090	CAACGCTACGAGGACGACGCGCAAGTTACTTGCCCATGCCAGGCGTCTACTTCTT	1149
Db	481		540
Qy	1150	CGCTTACAGTGTCTATGGCGCGCGCGACGCGACACGAGATGTGGGCCGACCTCATGAA	1209
Db	541		600
Qy	1210	GAACGGACAGTTCGGGCGCAGCGCCATTGCTCAGGACGCGGACGAGAACTACGACTACGC	1269
Db	601		660
Qy	1270	CAGCAACAGGCTTACTGCACTTGAAGTGGGGGAGAGGTCTTCAATCAAGTGGACGG	1329
Db	661		720
Qy	1330	CGGGAAAGTGCACGGCGCAACCAACAAGATACAGCACTTCTCCGGCTTCATCATCTA	1389


```

/db_xref="GI:10566471"
/translation="MWLLVILIPVLVSSAGTSAHYEMLGTGCMVCDPVGGTGKAPSTA
ATPDRGLMSLPTFIQGGEGARPGKAGRPGRPPGPPGVPGEGERGECROGLP
GPGPGLNAAGAISAISTVPTKATFYAGLRKHEGYEVLFDVVITNLGNHYPDTT
GKFTCSIPGIYPTTYHVLARMGGDGTSMWADLCKNNQVRSASIAQDADQNYDYASNSV
LHLEPGDEVYIKLDGSKAHGNKNKYSTFSGFIYAD"
BASE COUNT      629 a   653 c   664 g   580 t
ORIGIN
Query Match      15.2%; Score 305.8; DB 10; Length 2526;
Best Local Similarity 63.7%; Pred. No. 3.5e-34;
Matches 541; Conservative 0; Mismatches 257; Indels 51; Gaps 3;

Qy 602 AGTGAGGAAGATTGCTGGCCCTGGCAGCGTCCGGCTGAGCGCGCAGAGAGGGTGGCG 661
   || || || || || || || || || || || || || || || || || || || || ||
Db 820 AGACAAGAGATCGGCCCGCGCTGAGCCGCGCTCGTGGCAGCAGGAGGCGCTGATCGCC 879

Qy 662 GCGCGCGCGCTCGAGTGGCCATGGTCTGCTCTGCTGGTGGCCATCCCGTGCCTGGTG 721
   || || || || || || || || || || || || || || || || || || || || ||
Db 880 GCGCGGCGCTGGGGGTGGTATGGTGCTTCTCTGCTGATCTCATCCCGGTGCTGGTG 939

Qy 722 CACAGCTCCCGCGGGCCAGCGCACTACGAGATGCTGGGTGCTGCCGCACTGTGTGCGAC 781
   || || || || || || || || || || || || || || || || || || || || ||
Db 940 AGCTCGGCCGCGACGTCGGCTCACTACGAGATGCTGGGCACCTGCCGCATGTCTCGCAC 999

Qy 782 CCGCATGGGCCCGTGGCCCTGGTCCCGAGCGCGCGCTGC-----T 823
   || || || || || || || || || || || || || || || || || || || || ||
Db 1000 CCCTATGGGGGACCAAGGCTCCACGACCGCGCCCACTCCGGACCGGGGCTTATGCAG 1059

Qy 824 TCGGTGCCCGCCCTTCCGCCAGCGCGCAAGGGAGAGTGGCGCGCGCGGGAAGCAGGC 883
   || || || || || || || || || || || || || || || || || || || || ||
Db 1060 TCCTGCCCACTTTCATCCAGGTGCCAAGGCGAGCGCGGTAGCGCGGGAAGGACGCG 1119

Qy 884 CTCGGGGGGCC-----CCTGACCAACCAGGTCCAAAGAGGGCCCCCAGAGAACCC 934
   || || || || || || || || || || || || || || || || || || || || ||
Db 1120 CCGCTGGGCGCCCGGGAGAGCCCGACCGCGCGGCCCGTGGCGGCCCGCGCGAGAAG 1179

Qy 935 GGCAGGCCAGGCCCGCGGCCCTCCCGGTCCAGGTCCGGGCGGGGTGGCGCCGCTGCC 994
   || || || || || || || || || || || || || || || || || || || || ||
Db 1180 GGAGCGCGGTTCGCAAGGCTTTCGCCCGGCCCGCTGGGGCGCCGCGCTGAACGCGGCC 1239

Qy 995 G-----GCTAGCTGCCTCGCATATGCAAGTCTTCTTACGGCGGCGCTG 1030
   || || || || || || || || || || || || || || || || || || || || ||
Db 1240 GGGCTATCAGTCGACCCACCTATAGCAGGTGCCCAAGATCGCTTTACGCTGGCCTC 1299

Qy 1031 CGCGGGCCCGCAGAGGTTACGAGGTGCTCGCTTCGACGACGTGTGTGACCAACAGTGGCG 1090
   || || || || || || || || || || || || || || || || || || || || ||
Db 1300 AAACGGCAGCAAGAGGCTACGAGGTGCTCAAGTTCGACGACGTGTGTAACCAATCTCGGA 1359

Qy 1091 AACGCTTACGAGCAGCCAGCGCAAGTTTACTTGCCCATGCCAGCGCTCTACTTCTTC 1150
   || || || || || || || || || || || || || || || || || || || || ||
Db 1360 AACCACTATATCCCAACCGGGCAAGTTCACTGCTCCATACCGGTATCTACTTCTTC 1419

Qy 1151 GCTTACCACGTCTCATCGCGCGCGCAGCGCACCATGTGGCGCGACCTCATGAAG 1210
   || || || || || || || || || || || || || || || || || || || || ||
Db 1420 ACCTACACGCTCTGATGCGCGGAGGAGCGCACCATGATGTGGGTGATCTCTGCAG 1479

Qy 1211 AACGGACAGTCCGGCGACGCGCATGTCTCAGGACGCGGACCAAGCACTACGACTACGCC 1270
   || || || || || || || || || || || || || || || || || || || || ||
Db 1480 AACAAACAGGTGGGTGTAGTGAATTTGCTCAAGATGCTGATCAGAAATACGACTATGCC 1539

Qy 1271 AGCAACAGCGCTCATCTGCACCTGGAGGTGGGCGACGAGTCTTCATCAGCTGGACGCC 1330
   || || || || || || || || || || || || || || || || || || || || ||
Db 1540 AGTAACAGTGTGGTCTTCACTTGGAACACGAGAGATGAAGTCTATATCAAAATAGATGGT 1599

Qy 1331 GGGAAAGTGCAGCGCGCAACACCAAGTACAGCACCTTCTCCCGGCTTCATCATCTAC 1390
   || || || || || || || || || || || || || || || || || || || || ||
Db 1600 GGGAAAGCCACCGAGGAAACACAACAATACAGCACATCTCTGSAATTAATATTAT 1659

Qy 1391 CCGCACTGA 1399
   || || || ||
Db 1660 GCTCACTGA 1668
   || || || ||

```

[illegible]

Db 62 GCGCGGACGTCGCTCACTACGAGATGCTGGGACCTGGCGCATGTCTCGACCCCTA 121
Qy 787 TGGCGCCCGTGGCCCTGGTCCCGACGCGGCCCTGC-----TTCGGT 828
Db 122 TGGGGGACCAAGGCTCCAGCACCGCGGCACCTCGGACCGGGCCCTTATGCACTCCT 181
Qy 829 GCGCCCTTCCCGCAGCGGCAAGGAGAGAGGTGGCGCGCGCGGGAAGAGCGCCGCG 888
Db 182 GCGCACCTTTCATCAGGCTCCAAAGCGAGCGCGGTAGGCGGGGAAGCGCGCGCG 241
Qy 889 GGGGCGC-----CCTGGACCAACAGGTCCTCAAGAGGCGCCCGGAGAACCGCGAG 939
Db 242 TGGGCGCGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGAGA 301
Qy 940 GCCAGGCG 995
Db 302 GCGGGTGCAGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 361
Qy 996 -----GCTAGCTGCGCTCGCATGCTTCTTCTTCTTCTTCTTCTTCTTCTT 1035
Db 362 TATCAGTGCAGCCACCTATAGCAGGTCGCCCAAGATCGCTTTTACGCTGCGCTCAAACG 421
Qy 1036 GCGCGGACGAGGTTACGAGGTCGCGCTTCGACGAGCTGTGACCAACGTGGGCAACGC 1095
Db 422 GCAGCAGAGGCTACGAGGTCGCAAGTTCGACGAGCTGTGTAACCAATCTCGGAAACCA 481
Qy 1096 CTAGGAGCGACG 1155
Db 482 CTATGATCCCAACG 541
Qy 1156 CCAGCTGCTATCG 1215
Db 542 CCAGCTGCTATCG 601
Qy 1216 ACAGTCCG 1275
Db 602 CCAGTCCG 661
Qy 1276 CAGGCTATCTGACGCTGGAGTGGCGGAGGCTTTCATCATGAGTGGGCGCGGAA 1335
Db 662 CAGTGGTCTTCCAGTGGAAACGAGAGATGAGTCTATATCAAAATGATGGGAA 721
Qy 1336 AGTCAGCG 1395
Db 722 AGCCCGGAGGAAACAAACAAATACAGCACATCTCTGATTTATTTATGCTGA 781
Qy 1396 CTGA 1399
Db 782 CTGA 785

RESULT 12
BC022724 847 bp mRNA linear ROD 07-AUG-2002
LOCUS Mus musculus, clone IMAGE:4503736, mRNA, partial cds.
DEFINITION BC022724
ACCESSION BC022724
VERSION BC022724.1 GI:18490546
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 847)
Strausberg,R.
Direct Submission
Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdpaxil@stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
Series: IRAC Plate: 44 Row: f Column: 10.
Location/Qualifiers
1. 847
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4503736"
/tissue_type="Eye, retina, mouse strain C57BL/6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<1..364
/codon_start=2
/product="Unknown (protein for IMAGE:4503736)"
/protein_id="AAH22724.1"
/db_xref="GI:18490547"
/translation="PHEGYEVLFKFDVVNLNNGYDAASGRKTCNIPGTFFTYHVLV
RGDGTSMADLCKNGOVRAIAQADADONYASNSVILHLDAGDEVFINKLGGKAH
GGSNKKYSTFSGFIYSD"
BASE COUNT 197 a 269 c 209 g 172 t
ORIGIN
Query Match 13.8%; Score 278; DB 10; Length 847;
Best Local Similarity 83.0%; Pred. No. 3.7e-30;
Matches 317; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
Qy 1037 CCCACGAGGGTTACGAGGTGCTGGCTTCGACGACGCTGGTGGCAACGCGC 1096
Db 2 CCTCATGAGGTTACGAGGTGCTCAAGTTTGACGAGCTGGTCAACCACTAGGCAACAAC 61
Qy 1097 TACGAGGAGCGGCGGCAAGTTTACTTGGCCCATGCCAGGCTTACTTCTTCGCTTAC 1156
Db 62 TACGATGCGGCGGCGGCAAGTTTACATGCAACATCTCTGGCACCCTACTTTTACCTAC 121
Qy 1157 CACGTGCTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1216
Db 122 CAGGTCTCTATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 181
Qy 1217 CAGGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1276
Db 182 CAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241
Qy 1277 AGCGTCAATTCGACCTGGAGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1336
Db 242 AGCGTCAATTCGACCTGGAGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 301
Qy 1337 GTGCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1396
Db 302 GCACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 361
Qy 1397 TGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1418
Db 362 TGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 383

RESULT 13
AX463516 1742 bp DNA linear PAT 15-JUL-2002
LOCUS AX463516
DEFINITION Sequence 80 from Patent WO0248337.
ACCESSION AX463516
VERSION AX463516.1 GI:21886290
KEYWORDS human.
SOURCE

Search completed: June 6, 2003, 20:04:01
Job time : 5240 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2003, 16:23:28 ; Search time 1334 Seconds

(without alignments)
3403.320 Million cell updates/sec

Title: US-10-005-499-377

Perfect score: 2016

Sequence: 1 aggaaggttaggagcgga.....tgggctacagctgctg 2016

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	717	35.6	717	24	ABK35597
2	620	30.8	804	24	ABK35598
3	305.8	15.2	2526	22	AAH48068
c 4	304.6	15.1	639	22	AAH21084
5	303.2	15.0	768	24	ABA90355
c 6	286.4	14.2	584	24	ABQ45772
7	286.4	14.2	584	24	ABQ45773
8	272.4	13.5	864	24	AAI57577
9	269	13.3	466	24	ABL78241
					Gene encoding nove
					Gene encoding nove
					Murine HSP47 inter
					Parallel detection
					Human polynucleoti
					Oligonucleotide fo
					Oligonucleotide fo
					DNA encoding novel
					Human ovarian canc

10	262.8	13.0	584	24	ABQ45770	Oligonucleotide fo
c 11	262.8	13.0	584	24	ABQ45771	Oligonucleotide fo
12	259.2	12.9	981	24	ABK10666	Human CRF-like pro
c 13	107.6	5.3	505	24	ABQ29322	Oligonucleotide fo
14	107.6	5.3	505	24	ABQ29323	Oligonucleotide fo
c 15	101.4	5.0	505	24	ABQ29324	Oligonucleotide fo
16	101.4	5.0	505	24	ABQ29325	Oligonucleotide fo
c 17	100.5	5.0	1059	24	AB199588	Mouse ischaemic co
18	98.2	4.9	1136	24	ABK63764	Rat sequence diffe
c 19	95.2	4.7	15558	22	AAI39926	Genomic sequence #
20	95.2	4.7	15558	22	AAI39926	Human digestive sy
c 21	93.2	4.6	600	22	ABA59702	Human foetal liver
22	93.2	4.6	600	22	AAK07971	Human brain expres
c 23	93.2	4.6	600	22	AAK33841	Human bone marrow
24	93.2	4.6	600	22	AAI39567	Probe #8253 used t
c 25	93.2	4.6	600	24	ABSO8688	Human genome-deriv
26	91.8	4.6	987	22	ABF28263	Monkey NANGO 245 o
c 27	91.8	4.6	1388	22	AAI57973	Monkey NANGO 245 o
28	91	4.5	997	22	AAI57973	Human polynucleoti
c 29	91	4.5	1185	22	AAI57977	Human polynucleoti
30	90.4	4.5	1357	22	AAI28672	Human zacrpa4 codin
c 31	87.8	4.4	1338	22	AAI44999	Human secreted pro
32	87.8	4.4	1338	22	AAI45000	Human secreted pro
c 33	86.2	4.3	1338	22	AAI44970	Human TANGO 253 co
34	86.2	4.3	1338	22	AAI44998	Human secreted pro
c 35	85.6	4.2	114955	20	AAI53491	Human adenosine A1
36	85.4	4.2	728	22	AAI44994	Human secreted pro
c 37	85.4	4.2	728	22	AAI44994	Human secreted pro
38	85.4	4.2	728	22	AAI44996	Human adipocyte-sp
c 39	85	4.2	1347	20	AAI24684	cDNA clone encodin
40	85	4.2	1377	20	AAI87258	Human PRO344 nucle
c 41	85	4.2	1377	20	AAI80052	Human PRO344 prote
42	85	4.2	1377	21	AAI58626	Human PRO344 prote
c 43	85	4.2	1377	21	AAI01241	cDNA encoding nove
44	85	4.2	1377	21	AAI46907	Human PRO344 cDNA
c 45	85	4.2	1377	21	AAI46907	Human PRO344 cDNA

ALIGNMENTS

RESULT 1

ABK35597
ID ABK35597 standard; DNA; 717 BP.

XX AC ABK35597;

XX DT 08-MAY-2002 (first entry)

XX DE Gene encoding novel human secreted or membrane-associated protein #16.

XX KW Human; secreted protein; membrane-associated protein; hypertension;
XX KW inflammatory disorder; neurological disorder; haematopoietic disorder;
XX KW skeletal developmental disorder; growth abnormality; autoimmune disorder;
XX KW neurodegenerative disorder; nervous system disorder; bacterial infection;
XX KW peripheral myelinopathy; viral infection; cancer; obesity; diabetes;
XX KW hypotension; sexual development disorder; blood disorder; gene; ds.

OS Homo sapiens.

XX WO200204600-A2.

XX PD 17-JAN-2002.

XX PF 12-JUL-2001; 2001WO-US21985.

XX PR 12-JUL-2000; 2000US-218033P.

XX PR 21-AUG-2000; 2000US-226517P.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

XX PA (SMIK) SMITHKLINE BEECHAM PIC.

XX PA (GLAX) GLAXO GROUP LTD.

KW SS.
 XX Homo sapiens.
 OS WO200181363-A1.
 PN 01-NOV-2001.
 XX 26-APR-2001; 2001WO-US13360.
 XX 27-APR-2000; 2000US-199963P.
 PR 11-MAY-2000; 2000US-20336P.
 PR 25-MAY-2000; 2000US-207087P.
 PR 26-MAY-2000; 2000US-207546P.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX Agarwal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;
 PI Lai Y, Xie Q;
 XX WPI; 2002-041392/05.
 DR P-PSDB; ABB53290.
 XX Novel polypeptides and polynucleotides useful as a vaccine for
 PT preventing and treating diseases associated the polypeptide, e.g.
 PT Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility,
 PT asthma, amnesias
 XX Claim 2; Page 61; 116pp; English.
 PS The invention relates to an isolated polypeptide comprising a 277, 480,
 CC 583, 581, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026,
 CC 844, 782, 262, 394, 471, 485, 286, 533, 495, 350, 619, 490, 462, 255,
 CC 784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as
 CC given in the specification. The polypeptides, modulators of the
 CC polypeptides and antibodies against the polypeptides are useful for
 CC treating diseases such as neurological and psychiatric diseases
 CC including Alzheimer's, paraneuronal palsy, Huntington's disease,
 CC myotonic dystrophy, anorexia and depression; cardiovascular diseases
 CC including congestive heart failure, Hodgkin's disease and myocardial
 CC infarction; respiratory diseases including asthma, chronic obstructive
 CC pulmonary disease, cystic fibrosis and adult respiratory distress
 CC syndrome; liver diseases including hypercholesterolaemia, cirrhosis,
 CC viral and nonviral hepatitis, Type II diabetes mellitus, and impaired
 CC glucose tolerance; renal disease including renal failure, acute tubular
 CC necrosis and glomerulonephritis; skeletal muscle diseases including
 CC Eulenburg's disease, hypoglycaemia and obesity; gastrointestinal
 CC diseases including myotonia congenita and intestinal obstruction; lymph
 CC diseases including lymphagietasia; diseases of placenta including
 CC choriocarcinoma; diseases of testes including testicular cancer,
 CC male reproductive diseases including low testosterone and male
 CC infertility; and disease of pancreas including diabetic ketoacidosis,
 CC Type 1 and 2 diabetes and obesity. The present sequence encodes a
 CC polypeptide of the invention.
 XX Sequence 768 BP; 155 A; 259 C; 233 G; 121 T; 0 other;
 SQ

Query Match 15.08; Score 303.2; DB 24; Length 768;
 Best Local Similarity 66.38; Pred. No. 2e-46;
 Matches 509; Conservative 0; Mismatches 208; Indels 51; Gaps 3;

Qy 683 ATGGTGTGCTGCTGCTGGTGGCCATCCCGCTGCTGGTGCACAGCTCCCGCGGCCACGCG 742
 Db 1 ATGGTGTGCTGCTGCTGGTGGTATCTCTCCCGTGTGCTGGTGGTGGTGGTGGTGGTGG 60
 Qy 743 CACTACGAGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 790
 Db 61 CACTACGAGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 120
 Qy 791 -----CCCCGTGCCCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 844
 Db 121 CCCAGCACCGCTGCCACGCCGCGGCGGCTCATGCGATGCTCCCTGCGCCACCTTCATCCAG 180

Qy 845 GGCCTCAAGGAGAGAGTGGCGCGCGCGGGAAGCAGGCTTGGGGGGCCCCCTGGA--- 901
 Db 181 GGCCTCAAGGAGAGAGTGGCGCGCGCGGGAAGCAGGCTTGGGGGGCCCCCTGGA--- 240
 Qy 902 -----CCACCAGTCCAAGAGAGGCGCGCGCGGGAAGCAGGCTTGGGGGGCCCCCTGGA 955
 Db 241 CCGGGGCGACCCCGGCGCGCGGGAAGCAGGCTTGGGGGGCCCCCTGGA--- 300
 Qy 956 CCGGGGCGACCCCGGCGCGCGGGAAGCAGGCTTGGGGGGCCCCCTGGA--- 995
 Db 301 CTGGCGGCGCGCGCGGCGCGGGAAGCAGGCTTGGGGGGCCCCCTGGA--- 360
 Qy 996 ----GCTACGTCCTCGCATTTGCTTCTACGCGGCGCTTGGCGGCGCGCGCGCGCGCGCG 1051
 Db 361 TACAGCAGCGGCGCGCGGGAAGCAGGCTTCTACGCGGCGCTTCAAGCGCGCAGCATGAAGGCTAC 420
 Qy 1052 GAGTGTGCTGCTTTCGAGCAGGCTGCTGACCAACAGTGGGCAAGCTTACGAGCAGCAGCAGC 1111
 Db 421 GAGTGTGCTGCTTTCGAGCAGGCTGCTGACCAACAGTGGGCAAGCTTACGAGCAGCAGCAGC 480
 Qy 1112 GGCAAGTGTACTTGGCGCGCGGCGCGGGAAGCAGGCTTACGAGCAGCAGCAGCAGCAGC 1171
 Db 481 GGCAGGTCACCTGCTCCATCCCGGCGCATCTTCTACCTACGAGCAGCAGCAGCAGCAGC 540
 Qy 1172 GGCAGGTCACCTGCTCCATCCCGGCGCATCTTCTACCTACGAGCAGCAGCAGCAGCAGC 1231
 Db 541 GGAGGGGCGCGCGCGGGAAGCAGGCTTCTGCTGCAAAACACACCGAGTGGTGGTGGTGGT 600
 Qy 1232 GCATGCTCAGGAGCGCGCGGGAAGCAGGCTTCTGCTGCAAAACACACCGAGTGGTGGTGGT 1291
 Db 601 GCAATGCTCAGGAGCGCGCGGGAAGCAGGCTTCTGCTGCAAAACACACCGAGTGGTGGTGGT 660
 Qy 1292 CTGACGTCGGCGCGCGGGAAGCAGGCTTCTGCTGCAAAACACACCGAGTGGTGGTGGTGGT 1351
 Db 661 TTGAGCGCGGAGATGAAGTCTATCAATAGATGGCGGGAAGCAGGCTTCTGCTGCAAAAC 720
 Qy 1352 ACCAAGTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1399
 Db 721 AACCAAAATACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 768

RESULT 6
 ABQ45772/C
 ID ABQ45772 standard; DNA; 584 BP.
 XX AC ABQ45772;
 XX DT 12-JUL-2002 (first entry)
 XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 32363.
 XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX OS Homo sapiens.
 XX PN WO200218632-A2.
 XX PD 07-MAR-2002.
 XX PF 01-SEP-2001; 2001WO-EP10074.
 XX PR 01-SEP-2000; 2000DE-1043826.
 XX PR 05-SEP-2000; 2000DE-1044543.
 XX PA (EPIC-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 584 BP; 100 A; 45 C; 207 G; 232 T; 0 other;

Query Match 14.2%; Score 286.4; DB 24; Length 584;
Best Local Similarity 76.6%; Pred. No. 2.3e-43;
Matches 389; Conservative 0; Mismatches 116; Indels 3; Gaps 3;
QY 1215 GACAGTCCGGCCCGCCGCTTGTCTCAGGACGCGGACGAGTACGACTACGCGACA 1274
DB 507 GATAATCGAACCAACGCGCATCTACTCAAAACGCGAACCAGTACGCTACGCGACA 448
QY 1275 ACAGGCGTATCTGACCTGGAGCTGGGGGACGAGGTCTTCTATCAAGCTGGAGCGGGA 1334
DB 447 ACAAGGTCTATCTACACCTTAACGCTAAAGACGGAATCTTCAACAACTAAACGAGAA 388
QY 1335 AAGTCAGCGGGCGGACACCAACAGTACAGACCTCTCCGCGTCTATCATCTACCCG 1394
DB 387 AAATACACGACGACACCAACCAATACAAACACCTTCTCCGACCTTATCATCTACCCG 329
QY 1395 ACTGAGCGCGCGCGCGCGCGCTGCGCGCTTCTCTCCGCTTCTCTCAGCGACAC - 1452
DB 328 ACTAAACGACCGCGCGCGCGCTGACCGCGCTGCGCGCTTCTCTCCGATCTCACCACCT 269
QY 1453 TCGTGGCG 1512
DB 268 TCTTAACCG 209
QY 1513 CTTTCGCTTCCCGGAGCGCGCTTAATGGCGCACTCTTGGTCTCAAGGGTATAGTGG 1572
DB 208 CTTTCGCTTCCCGTAAACGACCTTAATAAACGACCTTAAATACCAAAATATAATAA 149
QY 1573 CCGGGAAGAGGAGGAGACCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1632
DB 148 CCGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 89
QY 1633 CCCAAGTGGCGCTGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1692
DB 88 CCCAATAAGCGCTAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1
QY 1693 TGGGAGGAGGTTTATAGCGGACTCA 1720
DB 28 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1

RESULT 7
ABQ45773

ID ABQ45773 standard; DNA; 584 BP.
XX
AC ABQ45773;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 32364.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 584 BP; 232 A; 207 C; 45 G; 100 T; 0 other;
Query Match 14.2%; Score 286.4; DB 24; Length 584;
Best Local Similarity 76.6%; Pred. No. 2.3e-43;
Matches 389; Conservative 0; Mismatches 116; Indels 3; Gaps 3;
QY 1215 GACAGTCCGGCCCGCCGCTTGTCTCAGGACGCGGACGAGTACGACTACGCGACA 1274
DB 78 GATAATCGAACCAACGCGCATCTACTCAAAACGCGAACCAGTACGACTACGCGACA 137
QY 1275 ACAGGCGTATCTGACCTGGAGCTGGGGGACGAGGTCTTCTATCAAGCTGGAGCGGGA 1334
DB 138 ACAAGGTCTATCTACACCTTAACGCTAAAGACGGAATCTTCAACAACTAAACGAGAA 197
QY 1335 AAGTCAGCGGGCGGACACCAACAGTACAGACCTTCTCCGCGTCTATCATCTACCCG 1394
XX

[illegible]

RESULT 11

RESOL 11
ABQ45771/c
ID ABQ45771 standard; DNA: 584 BP.

AC ABQ45771;

DT 12-JUL-2002 (first entry)

Human; cytosine methylation: 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; gastrointestinal; respiratory system; single nucleotide polymorphism; SNP; cell differentiation; ds.

OS Homo sapiens.

XX
XX
: curdson curdson

PN WO200218632-A2.

XX
XX
XXXXX

PD 07-MAR-2002.

XX

PF 01-SEP-2001;

XX

PR 01-SEP-2000;

PR 05-SEP-2000;

XX

PA (EPIG-) EPIG

XX 11

PI Olek A, Ple

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ13410-ABQ5054121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 584 BP: 262 A: 167 C: 45 G: 110 T: 0 other: 0

Query Match	13.0%	Score 262.8;	DB 24;	Length 584;
Best Local Similarity	72.4%;	Pred. No. 4.6e-39;		
Matches 368;	Conservative 0;	Mismatches 137;	Indels 3;	Gaps 2;

QY	1215	GACAGGTCCGGGCACGGCCATTGCTCAGGACGCGGACCAGAACTACGACTACGCCAGCA	1274
Db	507	GATAGGTCCGGGTTCAGCGTTATTTGTTTACGACGCGGATTAGAAATTCAGATTCAGTTAGTA	448

Accession	Sequence	Length
Qy	1275 ACACGGTCATCTGCACCTGGACGTGGCGCGACGAGGCTTTCATCAAGCTGGACGGCGGGA	1334
Db	447 ATACGGTATTTTGTATTTGGACGTGGCGCGACGAGTTTTTATTAAGTTGGACGGCGGGA	388

Qy	1335	AAGTGCACGGGGCACAACACCAAGTACAGCACCTTCTCCGGCTTCATCATCTACCCCG	1394
Db	387	AAGTGTACGGCGTAATATTAATAGTATAGTATTTTTTTCG - TTTATTATTTATTCG	329

[illegible]

Qy	1453	TCTGTCCCGCCACCGAGGCGCCACCCCTTTGAGAGCTGGCGGTGGGTGGAC	1512
Db	268	TTTTTGGTGGTATTATTCGAGGCGTTATTTTATTTTGGTGGAGCTTTGGCGTGGGTGGCAT	209

Qy	1513	CCTTCCGTTCCCGAGCGGGCCTAAATGGGGCAACTCTTGGTCTCAAGGGGTATAAGTGG	1572
Dh	208	TTTTTCCCTTTCGTTGTCGGGTTTAAATGCGGGCAATTTTTTGGTCTTTAAGGGTATAAGTGG	149

QY	1573	CGGGGAGAGGAGGACCCGCCAGAGGAGCAGAGCGACTTCCGGAGGGATCACCCGCA	1632
DB	148	TCCGGGACCGCGCGCGCTTCGGGTACGCGCATTTTCGGAGGGATTAATTTCGTA	89

[illegible]

QY TGGGGGACAGTTTAAATAGCGGACTCA 1720

28 TCGCCGCCTTCCTCCTGTTCTTCTCCGCCGTCTATT 1

RESULT 12

ABK10666

ID ABK10666 standard: CDNA: 981 BP.

XX
XX
XX

AC ABK

XX
XX

```
XX
PI      Olek A, Piepenbrock C, Berlin K, Guetig D:
XX      WPI; 2002-371829/40.
XX
PT      Determining the degree of cytosine methylation in genomic DNA, useful
PT      for diagnosis and prognosis, comprises selective hybridization of
PT      amplicons from chemically treated DNA
XX
XX      Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC      This invention describes a novel method for determining the degree of
CC      methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC      genomic sample of DNA. The sample is treated chemically to convert
CC      cytosine (C) but not methylated C, to uracil, then part of the genomic
CC      DNA that contains the target C is amplified to form a labeled amplicon.
CC      The amplicon is hybridised to two classes, each with at least one
CC      member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC      and the degree of hybridisation to both classes is determined from the
CC      label on the amplicon. From the ratio of labels hybridised to the two
CC      classes of oligomers, the degree of methylation is calculated. The method
CC      is used: (i) for diagnosis and/or prognosis of side effects of
CC      therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC      of the central nervous, cardiovascular, gastrointestinal and respiratory
CC      systems etc., particularly by detecting mutations or single nucleotide
CC      polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC      types and for investigating cell differentiation. The method allows the
CC      methylation status of many C residues to be determined simultaneously.
CC      ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC      method for determining the degree of cytosine methylation described in
CC      the disclosure of the invention.
XX
SQ      Sequence 505 BP; 79 A; 45 C; 123 G; 258 T; 0 other:
      Query Match          5.0%; Score 101.4; DB 24; Length 505;
      Best Local Similarity 72.1%; Pred. No. 1.2e-09;
      Matches 132; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY      1218 AGGTCCGGCCAGCGCCATTGCTCAGGACGCGGACCAACTACGACTACGCCAGCAACA 1277
DB      65 AGGTCCGGGTTAGTGTATTGTTTAGGACGCGGATTAGAATTACGATTACGTTAGTAATA 124
QY      1278 GCGTCATTCTGCACCTGGACGTGGCGGACGAGGCTCTTCATCAAGCTGGACGGCGGAAG 1337
DB      125 GCGTCATTCTGTATTGGACGTCGGCGACGAGGTTTTTATTAAAGTTGGATGGAGGTAAG 184
QY      1338 TGCACGGCGGCAACCAACAAGTACAGCACCTCTTCGCGCTTCATCATCTACCCGACT 1397
DB      185 TATACGGCGGTAAATAGTAATAATATAGTACGTTTTTTGGTTTTATTATTATTTCGATT 244
QY      1398 GAG 1400
DB      245 GAG 247
```

Search completed: June 6, 2003, 18:36:34
Job time : 1344 secs

Db	523	AGC-----CGGGTGCCTCCGCGCGTCTGACGACACCTTGGCCCTTCGACCGCGGTGCTGGTGA	577
Qy	1083	ACGTGGGCAACGCCTACGAGGACGCCAGCGGCAAGTTTACTTGGCCCATGCCAGGCGTCT	1142
Db	578	ACGAGCAGGACCATTACGACGCGCTCACCGGCAAGTTCACCTGCCAGGTGCTTGGGGTCT	637
Qy	1143	ACTTCTTTGCTTACCACGTGCTCATCGCGCGCGGC-GACGCGACCCAGCATGTGGCGCGAC	1201
Db	638	ACTACTTTCGCGCTCCATGCCACCGTCTACCGGCCAGCCTGCAGTTTGATCTGTTGAAGA	697
Qy	1202	CTCATGAAGAACGGACAGGTGCGGGCCAGCGCCATTCCTCAGGACGCGGACACAGACTAC	1261
Db	698	ATGCGCAATCCATTGCCTCTTCTTCCAGTTTTTCGGGGGTGGCCCAAGCCAGC-CTCG	756
Qy	1262	GACTACGCCAGCAACGCGTCATCTCGACCTGGACGTGGCGGACGAGGTCTTCATCAAG	1321
Db	757	CTCTCGGGGGGGCCATGGTAGGCTGGAGCCTGAGGACCAAGTGTGGGTGCAGGTGGGT	816
Qy	1322	CTGAGCGCGGGGAAAGTGCAACGGCGGCAACCAACAAAGTACAGCACCTTCTTCGGGCTTC	1381
Db	817	GTGGGTGACTACATTGGCATCTATGCTCAGCATCAAGACAGACGACCTTCTCCGGATT	876
Qy	1382	ATCATCTACCCCGACTGAGCGGGCGCGC	1410
Db	877	CTGGTGTACTCCGACTGGCACAGTCCCC	905

```

RESULT 2
US-09-140-804-1
: Sequence 1, Application US/09140804
: Patent No. 6197930
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.
: APPLICANT: Humes, Jacqueline M.
: TITLE OF INVENTION: ARIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
: FILE REFERENCE: 97-49
: CURRENT APPLICATION NUMBER: US/09/140, 804
: CURRENT FILING DATE: 1998-08-26
: EARLIER APPLICATION NUMBER: 60/056, 983
: EARLIER FILING DATE: 1997-08-26
: NUMBER OF SEQ ID NOS: 47
: SOFTWARE: FastSEQ for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 1347
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (198)...(926)
: US-09-140-804-1

```

	Query Match	4.28;	Score 85;	DB 4;	Length 1347;
	Best Local Similarity	49.1%;	Pred. No. 3.6e-09;		
	Matches 368;	Conservative 0;	Mismatches 370;	Indels 11;	Gaps 5;
Qy	663	GC	GGCGCGCGTGGAGTGC	GCATGCTGCTGCTGCTGGTGCCATCCCGCTGCTGTGTC	722
Db	178	GGGCTCCCGTGC	CAGCGCTATGAGGCACATCCCTCGTCTCTGCTGCTTGGGCTTGGCGG	237	
Qy	723	ACAGTCCCGCGGGCCACG	GCATACGAGATGCTGGGTGCTGGCGCGCATGCTGTGTCGAC	782	
Db	238	CCGCTCGCGCCCACTGG	AGCACAACAGATCCCCAGCCTCTGCTCCGGGGCACCCCGGCC	297	
Qy	783	CGCATGGGCCCCGTGG	CGCTGGTCCCGACGGCGGCGCTGCTTCGCTGCGCCCCCTTCCGCG	842	
Db	298	TTCCAGGACGCGCGGGC	----ACCATGGCAGCCAGGGCTTCCCGGGCGCGCATGGCGCGCG	354	
Qy	843	CAGCGCCCAAGGGAGTG	TGGCGCGCGGGAAGCAGAGCGCTCGCGGGGGCCCCCTCGAC	902	
Db	355	ACGGCCGCGACGCGCG	CCCCCGGGGCTCCGGGAGAGAAAGGCGAGGGCGGGAGGCGGGGAC	414	
Qy	903	CACCAAGTCCAAAGAG	GGCCCCCGCAGGAAACCCCGCAGCGCCAGGGCCCCCGGGGCCCTCCCG	962	

Db	415	TGCGGGACCTCGAGGGGACCCCGCGCGGGAGGAGCGGGACCCCGCGGGCCACCG	474
Qy	963	GTCACAGTCGCGGGGTGGCCCGCTGCCGGCTACGTGCGTTCGATTCGTTCTACG	1022
Db	475	GGCC-TGCGGGGAGTGCTCGGTGCTTCGCGCATCCGCTTCAGGCCCAAGCGCTCCGAG	533
Qy	1023	CGGGCTGCGCGCGGGCCCGCCACGAGGGTTACGAGGTGCTGCGCTTCGACGACGCTGGGTACCA	1082
Db	534	AGC-----CGGGTGCCCTCGCGCGCTCTGAGCGACCCCTTGCCCTTCGACCGCGTCTGGTGA	588
Qy	1083	ACGTGGGCAACGCGCTACGAGGCGACGAGCGGCAAGTTTACTTTGCCCCCATGCCAGGCGTCT	1142
Db	589	ACGAGCAGGGACATTAGCAGCGCGCTCACCGCAAGTTCACCTGSCCAGGTGCCTGGGGTCT	648
Qy	1143	ACTTCTTCGCTTACGACGTGCTCATGGCGGGCGGC--GACGGGACACGACATGTGGGCCGAC	1201
Db	649	ACTACTTTCGCGCTCCATTCGCCACCGTCTACCGGGCCAGCCTGTAGTTGTACTGGTGAAGA	708
Qy	1202	CTCATGAAGAACGACAGAGTGTCCGGGGCAGCGCCATTGCTCAGGACGCGGACACGAACTAC	1261
Db	709	ATGCGGAATCCATTGCGTCTTTCTTCAGTTTTTCGGGGGTGGGCCAAGCCAGC-CTCG	767
Qy	1262	GACTACGCCAGCAACACCGTCATCTGTCACCTGGACGTGGGCGCACGAGGTCTTCATCAAG	1321
Db	768	CTCTCGGGGGGGCCATTGGTGAGGCTTGAGCCCTGAGGACCAAGTGTGGGTGCGAGGTGGGT	827
Qy	1322	CTGAGCGCGGGGAAAGTGCACGGCGGCAACACCAAGTACAGCACTTCTCCGCGCTTC	1381
Db	828	GTGGGTGACATCATTTGCCATCTATGCGAGCATCAAGACAGACGACCTTCTCCGCGATT	887
Qy	1382	ATCATCTACCCGAGCTAGCGGGCGCGC	1410
Db	888	CTGGTGACTCCGACTGGACAGCTCCCG	916

RESULT 3
US-09-336-536-2
; Sequence 2, Application US/09336536
; Patent NO. 6406884
; GENERAL INFORMATION:
; APPLICANT: Leiby, K.
; APPLICANT: McKay, C.
; APPLICANT: Bossong, S.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-144
; CURRENT APPLICATION NUMBER: US/09/336,536
; CURRENT FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 728
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-336-536-2

Query Match	4.28;	Score 83.8;	DB 4;	Length 728;
Best Local Similarity	4.94;	Pred. No. 5.6e-09;		
Matches 360;	Conservative 0;	Mismatches 357;	Indels 12;	Gaps 5;
QY	683	ATGTTGCTGCTGCTGCTGGTGGCCATCCCGCTCTGCTGTCACAGCTCCGCGCGGCAGCG	742	
Db	1	ATGAGGCATCTCTTCGCTCTGCTCTCTGGGCTGGCGCGCTGCCCCACATGGAC	60	
QY	743	CACATCAGATCTGGTGTGCTGCGCATGTTGTGCAACCGCATGGCCCCGTGGCCCT	802	
Db	61	GACAAACAGATCCCCAGCCTCTGCCGGGGCACCCGGGCTTCACGACAGCCGGGCCAC	120	
QY	803	GGTCCGACGGCGGCGCTGCTTCGTTGCCCGCCCTTCCCGACGCGCCAGGAGAGGTG	862	
Db	121	CATGGCAGC-----CAGGGCTTGGCGGGCCGCGATGGCCGCGAGCGCCGACCGCGCGCC	176	
QY	863	GGCGGGCGGGAAAGCAGAGCCTTCGGGGGGCCCCCTTGACACACAGGTCCAGAGGGGCC	922	

```

177 CGGGGCTCCGGGAGAGAAAGCCGAGGGCGGGAGAGCGGGGACTGCCGGGACCTCGAGGGGAC 236
QY 923 CCAGGAGAAACCCCGGACAGCCAGGACCCCGCGGGCCCTCCCGGTCCAGGTCCGGCGGGGTG 982
Db 237 CCCGGGCGGAGGAGAGCGGACCCCGCGGGCCACCGGGCCTGCCGGGAGTGTCTC 295
QY 983 GCSCCCGCTCCCGGCTACGTCCTCGCATTTCTTCTACCGCGGGCCTTCGCGGGCCCCAC 1042
Db 296 GGTGCTCCGCGATCCGCTTACAGCGCCAAAGCGCTCCGAGAGC-----CGGGTGCCTCCG 350
QY 1043 GAGGGTTACGAGGTGCTCGCTTTCGACGACGTGTCACCAACGTCGGGCAACGCTACGAG 1102
Db 351 CGCTCTGACGCCACCTTGCCCTTCGACCGCGTGCTGTGAACGAGCAGGAGACATTTAGCAC 410
QY 1103 GCAGCCAGCGGCAAGTTTACTTCCCGCATGCCAGGGCTCTACTTCTTCGCTTACCACGTG 1162
Db 411 GCCTCACCGGCAAGTTTCACTGCCAGGTGCTTGGGTCTACTTACTTCGCGGTCCATGCC 470
QY 1163 CTATCGCGGGCGGC-GACGGCACCAACATGTGGGCGGACCTCATGAGAACCGACAGGT 1221
Db 471 ACCGTCTACCGGGCCAGCCTTGCAAGTTTGAATCTCGTGAAGAATGCGGAATCCATTGCTCT 530
QY 1222 CCGGGCCAGCGCCATTGCTCAGGACGCGGACCAAGAACTACGACTACGCCACGCAACAGCGT 1281
Db 531 TTCTTCCAGTTTTCGGGGGTGGCCCAAGCCAGC-CTCGCTCTCGGGGGGGGCCATGGT 589
QY 1282 CATCTCGACCTCGAGCTGGGCGACGAGTCTTCATCAAGCTGGACGGCGGGAAGTGCA 1341
Db 590 GAGGCTGGAGCCTGAGGACCAAGTGTGGGTGCAAGTGGGTGCTGGGTGACTTACATTTGGCAT 649
QY 1342 CGGCGGCAACACCAACAAGTACAGCACCTTCTCGGGTTCATCATCTACCCGACCTGACG 1401
Db 650 CTATGCGAGCATCAAGACAGACAGCACTTCTCCGATTTCTGGTGACTCCGACTGGCA 709
QY 1402 CGGCCCCCGC 1410
Db 710 CAGCTCCC 718

RESULT 4
US-09-140-804-10
; Sequence 10, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 729
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate nucleotide sequence encoding the zsig39
US-09-140-804-10

```

[illegible]

Db	62	AYAYAAARATHCCNWSNYTNTGYCCNGGNCAYCCNGGNTTNCNGGNACNCCNGGNCAYC	121
Qy	810	ACGGCGCGCTGCTT---CCGTGCCCCCTTCCGCCAGCGCGCAAGAGGAGAGGTGGGCC	866
Db	122	AYGGNWCNARGGNYTNCNGGNGMNGAYGGNMGNGAYGGNCCNCGGNG	181
Qy	867	GGCGCGGGAAGCAGCGCTCCGGGGGGCCCCCTTGACCAACAGGTCCAAGAGGCCGCCACG	926
Db	182	CNCCNGNGARAAAGGNGARGGNGGNGNCCNGGNTTNCNGGNCNMGNGNGAYCCNG	241
Qy	927	GAGAACCCGCGACGGCCAGGCCCCCGGGGCCCTCCCGGTCCAGTCCGGCGGGGTGGCGC	986
Db	242	GNCNMGNGNGARGCNGGNGGNCNACNGGNCNGGNGGNGARTGYWSNGTNC	301
Qy	987	CCGCTGCCGCTACGTGCCTCGATTGCTTCTACGGGGCGCTCGCGGCCGCCACGAGG	1046
Db	302	CNCCNMGWSNGCNTTWSNGCNARMGWSNGARWSNMNGTNCNCCNCCNWSNGA--	359
Qy	1047	GTTACGAGGTGTCGCGTTCGACGAGCTGTGTACCAACGCTGGGCAACGCTACGAGGCAG	1106
Db	360	-----YGCNCCNYTCCNTTYGAYMNGTNTNGTNAAYGARCARGGNCAYTAYGAYCNG	415
Qy	1107	CCAGCGCAAGTTTACTTGCCTCCCATCGCAGCGCTCTACTTCTTCGCTTACCAGTGCTCA	1166
Db	416	TNACNGNNAARTYACNTTYACGTTCARGTNCNGGNGTNTAYTAYTYGNGTNCAYGCNACNG	475
Qy	1167	TGCGCGGCGCGACGGCACGAGTGTGGCGCCACCTCATGAAGACGACAGGTCGCGG	1226
Db	476	TNTAYMNGCNWSNYTNCARTTYGAYTNGTNAARYGGNGARWSNATHGCNWSNTTYT	535
Qy	1227	CCAGCGCCATGCTCAGGACGGGACAGAACTACGACTACGCCACGACAGCGTCATTC	1286
Db	536	TYCARTTYTYGGNGGTGCCNNAARCCNCCNNSNTNWSNGGNGGNCNATGGTNMGNY	595
Qy	1287	TGCACCTGGACGTGGGCGACAGGTCTTCATCAAGCTGGACGCGGGAAGACGACGGCG	1346
Db	596	TNGARCCNGARGAYCARGTWGGTNCARTNGGNGTNGGNGAYTAYATHGGNATHTYAG	655
Qy	1347	GCAACACCAAGTACAGCACCTTCTCCGGCTTCATCATCTACCCCGACTG	1398
Db	656	CNWSNATHAARACNGAYWSNACNTTWSNGGNTTYTNGTNTAYWSNGAYTG	707

```

RESULT 5
US-09-118-408-1
; Sequence 1, Application US/09118408A
; Patent No. 6265544
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-30
; CURRENT APPLICATION NUMBER: US/09/118,408A
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 60/053,154
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2769
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (171)...(1013)
US-09-118-408-1

```

	Query Match	3.7%	Score 73.6;	DB 4;	Length 2769;
	Best Local Similarity	47.5%;	Pred. No. le-06;		
	Matches 291; Conservative	0;	Mismatches 309;	Indels 12;	Gaps 2;
Qy	851	AAGGGAGAGTGGGCGCGCGGGAAAGCAGGCTGC	GGGGGGCCCCCTGCCACCACCAAGT	910	
Db	462	AAAGGGGAGAAGGTGACCCGAGATCGAGGCTCC	AAGGGAANAATATGCAAAACAGGC	521	

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Arnold, White & Durkee
;; STREET: P.O. Box 4433
;; CITY: Houston
;; STATE: Texas
;; COUNTRY: United States
;; ZIP: 77210
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/843,659
;; FILING DATE: Concurrently Herewith
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Highlander, Steven L.
;; REGISTRATION NUMBER: 37,642
;; REFERENCE/DOCKET NUMBER: ARSB:519
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (512) 418-3000
;; TELEFAX: (512) 474-7577
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4257 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-843-659-1

Query Match 3.1%; Score 61.8; DB 4; Length 4257;
Best Local Similarity 47.6%; Pred. No. 0.00032;
Matches 245; Conservative 0; Mismatches 267; Indels 3; Gaps 2;
QY 467 CGCGCGCTGGAGCCGACAGCAGCAGGTAAGGGAGGGGGGAGCGCTCGGGGAGT 526
DB 2211 CGCGAGCGCGCCCGCTCGCGCGCTGCTGCGAGCTGCGTTCGTCGCGAGCGCT 2270
QY 527 GCAGACCCAGCGCCAGCGGGTACCGCTCTGCGCGCGG-AGAGCCCCCGCGCG 585
DB 2271 GGTGCTCATGCGCTCGCGGGGACCTGCGGTGCGCGCGGAGCGCGCGCTGCG 2330
QY 586 AGCCATTCGCCCAAGAGTGAAGATTGTTGCGCCCTGGCAGCGTCCGCGCTGAGCG 645
DB 2331 CGCGCTGCGCGCTGAGCTGCTGCGCGCGCGCTGGCGCGCGCTGCGCGGACCC 2390
QY 646 GCGCAAGAGGTGGCGGGCGCGCGCTCGGAGTGGCATGCTGCTGCTGCTGCTG 705
DB 2391 GCGCTGCGGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTG 2450
QY 706 CATCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 763
DB 2451 GCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2510
QY 764 TCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 823
DB 2511 CGCGCGCTCG 2570
QY 824 TCGGTGCCCCCTTCG 883
DB 2571 GCGCGAGCG 2630
QY 884 CTGCGGGGGCGCGCTGAGCAGCAGGTCCAGAGGGCGCGCGCGCGCGCGCGCGCG 943
DB 2631 CTCGAGCG 2690
QY 944 GCG 978
DB 2691 GCG 2725

RESULT 14

US-08-458-568A-11/c
;; Sequence 11, Application US/08458568A
;; Patent No. 5821339
;; GENERAL INFORMATION:
;; APPLICANT: Schaffer, Priscilla A.
;; APPLICANT: Yeh, Lily
;; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
;; TITLE OF INVENTION: Infections
;; NUMBER OF SEQUENCES: 15
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339r1s
;; STREET: One Liberty Place, 46th floor
;; CITY: Philadelphia
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19103
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: WordPerfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/458,568A
;; FILING DATE: 02-JUNE-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/065,146
;; FILING DATE: 05-MAY-1993
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Leary Ph.D., Kathryn R.
;; REGISTRATION NUMBER: 36,317
;; REFERENCE/DOCKET NUMBER: DFCI-0029
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 568-3100
;; TELEFAX: (215) 568-3439
;; INFORMATION FOR SEQ ID NO: 11:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 12001 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Herpes simplex virus
;; STRAIN: Herpes Simplex Virus Type 1
;; US-08-458-568A-11

Query Match 3.1%; Score 61.8; DB 1; Length 12001;
Best Local Similarity 47.6%; Pred. No. 0.0004;
Matches 245; Conservative 0; Mismatches 267; Indels 3; Gaps 2;
QY 467 CGCGCGCTGGAGCCGACAGCAGCAGGTAAGGGAGGGGGGAGCGCTCGGGGAGT 526
DB 4463 CGCGAGCGCGCGCGCGCTGCTGCGAGCTGCGGTGCTGCGCGCGCGCTGCTGCG 4404
QY 527 GCAGACCCAGCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 585
DB 4403 GGTGCTCATGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4344
QY 586 AGCCATTCGCCCAAGAGTGAAGATTGTTGCGCCCTGGCAGCGTCCGCGCTGAGCGG 645
DB 4343 CGCGGTGCGCGCTGAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4284
QY 646 GCGCAAGAGGTGGCGGGCGCGCGCGCTCGAGTGGCCATGCTGCTGCTGCTGCTGCTG 705
DB 4283 GCGCGCTGCGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4224
QY 706 CATCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 763
DB 4223 GCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4164

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2003, 20:04:08 ; Search time 308 Seconds

(without alignments)
9160.608 Million cell updates/sec

Title: US-10-005-499-377

Perfect score: 2016

Sequence: 1 aggaaggttagggagcgga.....tggggctacagctgctgctg 2016

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 870385 seqs, -699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCT05_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	269	13.3	466	10	US-09-867-701-1219
2	258.8	12.8	795	9	US-09-998-563-1
3	237.8	11.8	750	9	US-09-998-563-3
4	181.4	9.0	804	9	US-09-998-582-1
5	176.4	8.8	804	9	US-09-998-582-3
6	98.2	4.9	1136	10	US-09-917-800A-1671
7	95.2	4.7	15558	9	US-09-764-872-823
8	93.2	4.6	600	10	US-09-864-761-11941
9	91	4.5	1185	9	US-10-098-841-182
10	88.2	4.4	1377	12	US-10-044-090-60
11	86.2	4.3	1338	9	US-09-796-753-67
12	85	4.2	1377	9	US-09-944-413-41
13	85	4.2	1377	9	US-09-944-403-41
14	85	4.2	1377	9	US-09-944-896-41
15	85	4.2	1377	9	US-09-944-944-41
16	85	4.2	1377	9	US-09-944-907-41
17	85	4.2	1377	9	US-09-944-929-41
18	85	4.2	1377	9	US-10-028-072-361
19	85	4.2	1377	9	US-10-121-049-361

20 4.2 1377 9 US-10-123-904-361 Sequence 361, App

21 85 4.2 1377 9 US-10-140-470-361 Sequence 361, App

22 85 4.2 1377 9 US-10-175-746-361 Sequence 361, App

23 85 4.2 1377 9 US-10-176-918-361 Sequence 361, App

24 85 4.2 1377 9 US-10-176-921-361 Sequence 361, App

25 85 4.2 1377 9 US-10-137-865-361 Sequence 361, App

26 85 4.2 1377 9 US-10-140-474-361 Sequence 361, App

27 85 4.2 1377 9 US-10-142-431-361 Sequence 361, App

28 85 4.2 1377 9 US-10-143-114-361 Sequence 361, App

29 85 4.2 1377 9 US-10-140-002-361 Sequence 361, App

30 85 4.2 1377 9 US-10-142-419-361 Sequence 361, App

31 85 4.2 1377 9 US-10-123-262-361 Sequence 361, App

32 85 4.2 1377 9 US-10-142-423-361 Sequence 361, App

33 85 4.2 1377 9 US-10-121-050-361 Sequence 361, App

34 85 4.2 1377 9 US-10-141-755-361 Sequence 361, App

35 85 4.2 1377 9 US-10-143-032-361 Sequence 361, App

36 85 4.2 1377 9 US-10-123-108-361 Sequence 361, App

37 85 4.2 1377 9 US-10-123-236-361 Sequence 361, App

38 85 4.2 1377 9 US-10-123-261-361 Sequence 361, App

39 85 4.2 1377 9 US-10-140-921-361 Sequence 361, App

40 85 4.2 1377 9 US-10-140-928-361 Sequence 361, App

41 85 4.2 1377 9 US-10-121-045-361 Sequence 361, App

42 85 4.2 1377 9 US-10-123-292-361 Sequence 361, App

43 85 4.2 1377 9 US-10-123-903-361 Sequence 361, App

44 85 4.2 1377 9 US-10-124-819-361 Sequence 361, App

45 85 4.2 1377 9 US-10-124-822-361 Sequence 361, App

ALIGNMENTS

RESULT 1

US-09-867-701-1219

; Sequence 1219, Application US/09867701

; Patent No. US20020132237A1

; GENERAL INFORMATION:

; APPLICANT: Aglate, Paul A.

; APPLICANT: Jones, Robert

; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER

; FILE REFERENCE: 210121.497

; CURRENT APPLICATION NUMBER: US/09/867,701

; CURRENT FILING DATE: 2001-05-29

; NUMBER OF SEQ ID NOS: 10912

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1219

; LENGTH: 466

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)....(466)

; OTHER INFORMATION: n = A,T,C or G

US-09-867-701-1219

Query Match 13.3%; Score 269; DB 10; Length 466;

Best Local Similarity 96.8%; Pred No. 6.8e-61;

Matches 338; Conservative 0; Mismatches 5; Indels 6; Gaps 6;

QY 1606 GAGCGACTTCGGAGGGATACCCGACCCAGTCGCGCTGGACCCCATAGGGCAGAG 1665

Db 1 GAGCGACTTCGGGA-GGATACCCGACCCCAAGTG-GCGCTGGACCCATA-GGGCAGAG 57

QY 1666 GTCGTGGCTTCTCTTTTGTACAGAGATGGGGAGGAGTTTAAATAGCGGGACTCAGAGC 1725

Db 58 GTCGTGGCTTCTCTTTTGTACAGAGATGGGGAGGAGTTTAAATAGCGGGACTCAGAGC 117

QY 1726 CCAGAAAGCCGGAGGAGCCCGCAGCTTGCAGGGGAAATAACAGAAACAGAGGAGAGC 1785

Db 118 CCAGAAAGCCGGAGGAGCCCGCAGCTTGCAGGGGAAATAACAGAAACAGAGGAGAGC 177

QY 1786 CCATTTAGGCAAGAGAGACATTAACAGAGGTAGTCAGGTTTCTCCGTCACAACTTTCT 1845

Db 178 CCATTAGCAAGAGACATTAACAGGAGGTAGTCAGGTTCTCCGTCACAACCTTCT 237
QY 1846 CTCGCCACCTCTCGTCCCTCGTCTCCACTTTTCAGGCTCAGGCTCCAGCCTTGCGAGCC 1905
Db 238 CTCGCCACCTCTCGTCCCTCGTCTCCACTTTTCAGGCTCAGGCTCCAG-CTTGCGAGCC 296
QY 1906 TTCCTGTGAACCT-GGAGGAACCACT-GAATTCCTTCTTCCCTGGCATTTAAAA 1952
Db 297 TTCCTGTGAACCTGGGAGGAACCACTGGAATTCCTTCTTCCCTGGCATTTAA 345
RESULT 2
US-09-998-563-1
; Sequence 1, Application US/09998563
; Patent No. US20020155546A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
; FILE REFERENCE: ZACRP12
; CURRENT APPLICATION NUMBER: US/09/998,563
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 60/253,867
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 795
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (46)...(795)
US-09-998-563-1

Query Match 12.8%; Score 258.8; DB 9; Length 795;
Best Local Similarity 70.9%; Pred. No. 3.4e-58;
Matches 343; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 912 CAAGAGGGCCCCAGGAGAACCCGCGCAGCCAGGCCCCCGGGCCCTCCGGTCCAGGTC 971
Db 311 CCAGCGGCTCGGGGTGGTGGCGCGCGGGCGGGGTAGGTGGCGATTCCGAGGGTGAAG 370
QY 972 CGGGGGGGTGGCGCCCTGCGGCTACGTCGCTGCGATTGCTTCTACGGCGGCTGC 1031
Db 371 TGACCACTGCGTGAAGCCACCTTCAGCGGCCCCAAGATCGCTTCTATGTGGTCTCA 430
QY 1032 GGGGGCCCCACAGAGGTTACGAGGTGCTGCGCTTCGACGACGTGTGACCAACGTTGGCA 1091
Db 431 AGAGCCCCCAGCAAGGCTATGAGGTGCTGAAGTTCGATGACGTGGTCAACCACTCGGCA 490
QY 1092 ACGCTACAGGACCCAGCGCAAGTTTACTTGCCCATGCGCGCTCTACTTCTTCG 1151
Db 491 ATCACTATACCCACCGGCAAGTTTCAGCTGCGAGGTACGCGGCATCTACTTCTCA 550
QY 1152 CTTACACCTGCTCATGCGCGGCGCGGACCGACAGCATGTGGCGCGACCTCATGAAGA 1211
Db 551 CTTACACATCTCATGCGCGGCGGCGGACCGACAGCATGTGGGGGACCTCTGCAAGA 610
QY 1212 ACGGACAGTCCGGGCGGCGGCGGCTGCTCAGGACGCGGACAGCACTACGACTAGGCA 1271
Db 611 ACGGCGAGTCCGGGCGGCGGCGGCTTGCACAGAGCGGCGGACAGCACTACGACTAGGCA 670
QY 1272 GCAACAGCTCATCTGCACTGAGGTGGGCGGAGGCTTTCATCAAGCTGGAGCGG 1331
Db 671 GTAACAGCTGCTGCTGCACTTGGATTCAGGGGACGAAGTGTATGTAAGCTGGATGGG 730
QY 1332 GGAAAGTGCAGCGGCGGCAACCAACAGTACAGCACTTCTCCGGCTTCATCATCTACC 1391
Db 731 GGAAGGCTCAGCGAGGCAATAACAGTACAGCACTTCTCCGGCTTCATCATCTACC 790
QY 1392 CCGA 1395

Db 791 CGGA 794
RESULT 3
US-09-998-563-3
; Sequence 3, Application US/09998563
; Patent No. US20020155546A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
; FILE REFERENCE: ZACRP12
; CURRENT APPLICATION NUMBER: US/09/998,563
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 60/253,867
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 750
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide
; OTHER INFORMATION: of SEQ ID NO:2
; NAME/KEY: variation
; LOCATION: (1)...(750)
; OTHER INFORMATION: Each n is independently T, A, G, or C
; NAME/KEY: misc_feature
; LOCATION: (1)...(750)
; OTHER INFORMATION: n = A,T,C or G
US-09-998-563-3

Query Match 11.8%; Score 237.8; DB 9; Length 750;
Best Local Similarity 52.0%; Pred. No. 1e-52;
Matches 213; Conservative 85; Mismatches 112; Indels 0; Gaps 0;
QY 987 CGGCTCGCGGCTACGTCGCTGCTGCTTCTACGCGGCTGCGGCGGCCCGGCGAGG 1046
Db 341 SNGCNACNTTYSNGNCCNAAATATGCTTTTATGCTTNGTNGTNGTNGTNGTNGTNG 400
QY 1047 GTTACGAGTGTGCGCTTCGACGAGTGTGTCACCAACCTGGGCAACGCTACGAGGAG 1106
Db 401 GNTAYGARTNTYNAARTTYGAYGTGTGACNAAYTNGNAAAYTAYGAYCCNA 460
QY 1107 CCAGCGGCAAGTTTACTTCCCGCATCCAGGCGTCTACTTCTCGCTTACCAGCTCA 1166
Db 461 CNACNGNAAATTYWSNTGYCARGTNMNGGNATHATYTTTAYACNTAYATHTYTA 520
QY 1167 TCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1226
Db 521 TCMGNGNGNGAYGNGNACNWSNATGTGGGCGNGAYTNTGYAARAYGNCARGTNMNG 580
QY 1227 CCAGCGCATCTGCTCAGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1286
Db 581 CNWSNGCNATHGNCARGAYGNGAYCARAAYTAYGAYTAYGAYGNGNAAAYWSNGTNG 640
QY 1287 TGCACCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1346
Db 641 TNCATYTTGAYWSNGNGAYGARGNTATGTNAAATYNGAYGNGNGNAAARGNCATYGG 700
QY 1347 GCAACACCAACAGTACAGCACTTCTCCGGCTTCATCATCTACCCCGAC 1396
Db 701 GNAAYAAAYAAAYTAYWSNACNTTYSNGGNTTYYTNTYNTAYCNGAY 750

RESULT 4
US-09-998-582-1
; Sequence 1, Application US/09998582
; Patent No. US20020160474A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian

Query Match	4.3%	Score 86.2	DB 9	Length 1338	
Best Local Similarity	49.3%	Pred. No. 5e-13			
Matches 369	Conservative	0	Mismatches 368	Indels 12	Gaps 5
QY	663	GC CGCGCGCTCGAGTGGCCATGGTGTGCTGTGCTGTGGTGTGGCCATCCCGCTGCTGTGGTGC	722		
Db	168	GGGCTCCCGGTGCGCAGCGGTATGAGGCCAATCTCGTCTCTGCTGTGCTCTCTGGGCTGTGGCG	227		
QY	723	ACAGCTCCCGCGCGGCAGCGCACTACGAGATGCTGGGTGCTGCTCCCGCATGGTGTGGCAAC	782		
Db	228	CCGGCTGCGCCCCACTTGGAGCAACAAGATCCCAAGCTTCCCGCGGGGCACCCCGGCC	287		
QY	783	CGCATGGGCGCGCTGCGCTGTGTCGCGACGGCGCGCTGCTTCGTCGCCCGCTTCCCGC	842		
Db	288	TTCAGCGACGCGCGGCCACCATGGCAGC-----CAGGGCTTGGCGGGCGCGATGGCGCG	343		
QY	843	CAGCGCCCAAGGAGAGGTGGCGCGCGCGGAAAGCAGGCCCTCGCGGGGCGCCCTTGGAC	902		
Db	344	GACGCGCGGACGCGCGCGCCCGGGGTCCGGGAGAGAAAGCGAGGCGGGAGCGGGAC	403		
QY	903	CACAGGTCAAGAGGGCCCCAGGAGAACCCCGCAGGCCAGGCCCGCGGGCGCTCCCG	962		
Db	404	TGCCGGGACCTCGAGGGGACCCCGCGCGGAGAGAGCGGGACCCCGCGGGGCCACCG	463		
QY	963	GTCCAGGTTCGGCGGGGTGGCGCGCGCTGCGCGGCTACGTGCTCGCATTCCTTTACG	1022		


```

Db      444  TGCCGGGACCTCGAGGGGACCCCGGCGCGAGGAGCGGACCCCGGGGCCACCG 503
QY      963  GTCCAGGTCGGGGGGGTGGCGCGCTGCGGCTACGCTGCTCCATTGCTTCTAGG 1022
Db      504  GGCC-TGCCGGGGAGTGTGCTGCTCCGATCCGCTTCAGGCGCAAGCGCTCCGAG 562
QY      1023  CGGGCCTCGGGGGGCCCGCCACGAGGTTACGAGGTGCTGGGCTTCGACGACGTGGTGACCA 1082
Db      563  AGC-----CGGTGCTCGCGCTGACGCACTTGGCCCTTCGACCGCGTGTGGTGA 617
QY      1083  ACGTGGCAACGCCCTACGAGGAGCGAGCGCAAGTTTACTTGGCCCATGCGCAGGCTCT 1142
Db      618  ACGAGCAGGGACATTACGACGCGCTACCGGCAAGTTACCTGCCAGGTGCTGGGCTCT 677
QY      1143  ACTTCTTCCTTACCAGTGTCTCATGCGGCGCGGC-GACGGCACCAGCATGTGGCCGAC 1201
Db      678  ACTACTTCGCCCTCCATGCCACCGTCTACCGGGCCAGCTGAGTTGATCTGGTGAAGA 737
QY      1202  CTCATGAAGAAGCGGACAGGTCCGGGCCAGCGCATTTGCTCAGGAGCGGACCAAGAACTAC 1261
Db      738  ATGGCGAATCCATTGCTCTTCTTCCAGTTTTCGGGGGGTGGCCCAAGCCAGC-CTCG 796
QY      1262  GACTAGCCGAGCAACAGCGTCTATTCGACCTGGAGCTGGGGCGGACGAGGTCTTCATCAAG 1321
Db      797  CTCTCGGGGGGGCCATGTTGAGGCTGGAGCTGAGGACCAAGTGTGGTGCAGGTGGGT 856
QY      1322  CTGGAGCGGGGAAAGTGCACGCGGCAACACCAAGTACAGACACCTTCTCCGGCTTC 1381
Db      857  GTGGGTGACTACATTGGGCACTATGCGGACATCAAGACAGACAGACCTTCTCCGGATT 916
QY      1382  ATCATCTACCCGAGTGGCGGCGCCCGC 1410
Db      917  CTGGTGTACTCCGACTGGCACAGCTCCCC 945

```

Search completed: June 7, 2003, 01:08:04
Job time : 315 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2003, 17:46:03 ; Search time 2931 Seconds
(without alignments)
11139.577 Million cell updates/sec

Title: US-10-005-499-377
Perfect score: 2016
Sequence: 1 aggaaggttagggagcgga.....tggggctacagctgctgctg 2016

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:**
2: em_esthum:**
3: em_estin:**
4: em_estmu:**
5: em_estov:**
6: em_estpl:**
7: em_estro:**
8: em_htc:**
9: gb_est1:**
10: gb_est2:**
11: gb_htc:**
12: gb_est3:**
13: gb_est4:**
14: gb_est5:**
15: em_estfun:**
16: em_estom:**
17: gb_gss:**
18: em_gss_hum:**
19: em_gss_inv:**
20: em_gss_pln:**
21: em_gss_vrt:**
22: em_gss_fun:**
23: em_gss_mam:**
24: em_gss_mus:**
25: em_gss_Other:**
26: em_gss_pro:**
27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	958	47.5	967	9	AL561826
C 2	874.8	43.4	914	9	AL529897
C 3	869.8	43.1	920	9	AL519829
4	868.6	43.1	942	9	AL529898
5	843.2	41.8	894	9	AL518440
C 6	824.8	40.9	870	9	AL518439

7	479.8	23.8	557	14	BM762336
8	355.8	17.6	387	12	BE888168
9	353.8	17.5	425	14	BM844470
10	327.8	16.3	845	13	BI603927
11	314.4	15.6	542	13	BM141973
12	314.4	15.6	552	13	BI793113
C 13	313.4	15.5	323	17	AQ200834
14	308.8	15.3	461	13	BI793031
15	307.6	15.3	540	13	BM354017
16	304.2	15.1	638	12	BF686380
17	297.8	14.8	301	14	BM847021
C 18	295.2	14.6	389	17	AZ793063
19	288.2	14.3	982	12	EG169143
C 20	286.2	14.2	307	17	AQ586940
C 21	269.8	13.4	556	13	BM128150
22	269	13.3	466	14	H61343
23	268.6	13.3	533	14	HQ101126
24	268.6	13.3	578	14	HQ101250
25	266	13.0	535	14	HQ101230
26	261.4	13.0	644	14	BM711545
27	259.4	12.9	491	13	BM128991
28	257.4	12.8	523	13	BM128566
C 29	256.2	12.7	654	13	BJ092454
C 30	255.6	12.7	334	12	BF593076
31	255.4	12.7	531	12	EG294846
C 32	240.6	11.9	885	14	HQ171748
C 33	236	11.7	782	9	AL567669
34	230	11.4	929	13	BI737761
35	221	11.0	842	12	EG325001
C 36	219.8	10.9	815	9	AL566412
37	219.4	10.8	297	10	BE650685
38	217	10.8	815	12	BF786942
C 39	198.2	9.8	518	12	BE944830
40	197.2	9.8	518	12	BF471505
41	188.6	9.4	383	13	BI104863
C 42	188.4	9.3	335	13	BI042421
43	186.4	9.2	318	13	BI042438
C 44	184.2	9.1	487	12	EG374818
C 45	181.6	9.0	391	9	AI683209

ALIGNMENTS

RESULT 1
AL561826/C
LOCUS AL561826 LTI_NFL004_NBC2 Homo sapiens cDNA clone EST 16-FEB-2001
DEFINITION prime, mRNA sequence.
ACCESSION AL561826
VERSION AL561826.1 GI:12909642
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 967)
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Location/Qualifiers
1..967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB005YK05"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com>"

172 a	276 c	317 g	200 t	2 others
-------	-------	-------	-------	----------

BASE COUNT
ORIGIN

Query Match	47.5%	Score 958;	DB 9;	Length 967;
Best Local Similarity	99.7%;	Pred. No. 7.7e-181;		
Matches 959;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1055	GTGCTGCGCTTCACGACAGCTGGTGACCAAGCTGGGCAACGCTTACGAGGCGAGCCAGCGGC	1114	
DB	963	GGTBTGCGCTTCAGCGACGCTGGTGACCAAGCTGGGCAACGCTTACGAGGCGAGCCAGCGGC	904	
QY	1115	AAGTTTACTTGGCCCCATGCCAGGCGGCTACTTCTTCGCTTACACGTGCTCATGCGCGGC	1174	
DB	903	AAGTTTACTTGGCCCCATGCCAGGCGGCTACTTCTTCGCTTACCCAGCTGCTCATGCGCGGC	844	
QY	1175	GGCAGCGGCACACGATGTGGCGCGACCTCATCAAGAACGGACAGTCCGGGCGACGCGC	1234	
DB	843	GGCAGCGGCACACGATGTGGCGCGACCTCATGAAGAACGGACAGTCCGGGCGACGCGC	784	
QY	1235	ATTGCTCAGGACCGGACCAAGAACTACGACTAGCGGACGCAACACGCTCATTTGCAACCTG	1294	
DB	783	ATTGCTCAGGACCGGACCAAGAACTACGACTACGCGAGCAACACGCTCATTTGCAACCTG	724	
QY	1295	GAGCTGGGCGACAGGCTTTCATCAGCTGGAGCGGGGGGAAAGTGCACGCGCGCACACC	1354	
DB	723	GAGCTGGGCGACAGGCTTTCATCAGCTGGAGCGGGGGGAAAGTGCACGCGCGCACACC	664	
QY	1355	AACAAGTACAGCACCTTCTCCGGCTTCATCATCTACCCGACGTAGCGCGGCCCGGCCCGC	1414	
DB	663	AACAAGTACAGCACCTTCTCCGGCTTCATCATCTACCCGACGTAGCGCGGCCCGGCCCGC	604	
QY	1415	TGCCCCCGCTCGGCCCTTCTTCGCTGCTTCACCCACCTCTCGCGCGGCCCGCCAGGC	1474	
DB	603	TGCCCCCGCTCGGCCCTTCTTCGCTGCTTCACCCACCTCTCGCGCGGCCCGCCAGGC	544	
QY	1475	GCCACCCACCTTTGAGAGCCTTGGCGGTGGGGTGAGACCTTCGCTTCCCGGAGGCGGCC	1534	
DB	543	GCCACCCACCTTTGAGAGCCTTGGCGGTGGGGTGAGACCTTCGCTTCCCGGAGGCGGCC	484	
QY	1535	TAAATGGCGCACTTTGTGCTCAAGGGTATTAAGTGGCGCGGAAGAGGAGAGACCCCGG	1594	
DB	483	TAAATGGCGCACTTTGTGCTCAAGGGTATTAAGTGGCGCGGAAGAGGAGAGACCCCGG	424	
QY	1595	CCAGAGGACAGAGCGACTTCCGAGGAGATCACCCGACCCCAAGTGGCGCTCGACCCCA	1654	
DB	423	CCAGAGGACAGAGCGACTTCCGAGGAGATCACCCGACCCCAAGTGGCGCTCGACCCCA	364	
QY	1655	TAGGGCGAGAGTCTGGGCTTCTCTTTGTACAGAGATGGGGAGCAGTTTTTAATAGCGG	1714	
DB	363	TAGGGCGAGAGTCTGGGCTTCTCTTTGTACAGAGATGGGGAGCAGTTTTTAATAGCGG	304	
QY	1715	GACTCAGAGCCACAGAAAGCCGAGGAGAGCCCCCGAGCTTCGGAGGGGAAATAACAGAA	1774	
DB	303	GACTCAGAGCCACAGAAAGCCGAGGAGAGCCCCCGAGCTTCGGAGGGGAAATAACAGAA	244	
QY	1775	ACAGGAGAGCCCATTTAGGCAAGAGAAACATTTAAACAGAGGTAGTGCAGGTTCTCCCGT	1834	
DB	243	ACAGGAGAGCCCATTTAGGCAAGAGAAACATTTAAACAGAGGTAGTGCAGGTTCTCCCGT	184	
QY	1835	CACAACCTTCTTCGCCACCCCTCTCGTCCCTCGTCTCCACTTTTCAGGCTCAGGCTCCAG	1894	
DB	183	CACAACCTTCTTCGCCACCCCTCTCGTCCCTCGTCTCCACTTTTCAGGCTCAGGCTCCAG	124	

Qy	1895	CCTTGGCAGCGCTTCTCTGTGAACCTGGAGAACCAACAGTGAATTCCTTCTGGCATTTAAACG	1955
Db	123	CCTTGGCAGCGCTTCTCTGTGAACCTGGAGAACCAACAGTGAATTCCTTCTGGCATTTAAACG	64
Qy	1955	CATTCTGTACAGTCCCAATTCCTCCCTATCCGGACTAGGCCCTCGGGCTACAGCTGCTGC	2014
Db	63	CATTCTGTACAGTCCCAATTCCTCCCTATCCGGACTAGGCCCTCGGGCTACAGCTGCTGC	4
Qy	2015	TG 106	
Db	3	TG 2	
RESULT 2	AL529897/c		
LOCUS	AL529897	LTI_NFL001_NBC4	914 bp mRNA linear EST 13-FEB-2001
DEFINITION	AL529897	Homo sapiens	CDNA clone CSDD005YH17 3
ACCESSION	AL529897	primer, mRNA sequence.	
VERSION	AL529897		
KEYWORDS	AL529897.1	GI:12793390	
SOURCE	EST.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 914)		
JOURNAL	Li, W.B., Gruber, C., Jesse, J., and Polayes, D.		
COMMENT	Full-length cDNA libraries and normalization		
Unpublished (2001)			
Contact: Genoscope			
Genoscope - Centre National de Sequencage			
BP 191 91006 EVRY cedex - France			
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.			
Location/Qualifiers			
1. 914			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="CSDD005YH17"			
/clone_lib="LTI_NFL001_NBC4"			
/sex="male"			
/tissue_type="neuroblastoma cells"			
/lab_host="DH10B"			
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA			
was primed with a NotI-oligo(dT) primer. Five prime end			
enriched, double-stranded cDNA was digested with Not I and			
cloned into the Not I and Eco RV sites of the pCMVSPORT 6			
vector. Library was normalized. Library was constructed			
by Life Technologies. Contact : Feng Liang Life			
Technologies, a division of Invitrogen 9800 Medical Center			
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610			
8371 Email : fliang@lifetech.com URL :			
http://fulllength.invitrogen.com"			
BASE COUNT	159 a	252 c	296 g 186 t 21 others
ORIGIN			
Query Match	43.4%	Score 874.8;	DB 9; Length 914;
Best Local Similarity	96.28;	Pred. No. 2.9e-164;	
Matches 870;	Conservative 21;	Mismatches 13;	Indels 0; Gaps 0;
Qy	1076	GTGACCAACGTGGGCAACGCCTACGAGGACGCCAGCGGCAAGTTTACTTCCCGCCATGCCA	1135
Db	904	GTGACCAACGTGGGCAACGCCTACGAGGACGCCAGCGGCAAGTTTACTTCCCGCCATGCCA	845
Qy	1136	GGCGTCTACTTCTTCGCTTACACAGTGCTCATCGGGGGGGCGACGCCACCATGTGG	1195
Db	844	GGCGTCTACTTCTTCGCTTACACAGTGCTCATCGGGGGGGCGACGCCACCATGTGG	785
Qy	1196	GCGGACCTCATGAAGAACGACAGGTCCGGGGCGAGGCCATTCTCTCAGGACGCGGACCG	1255
Db	784	GCGGACCTCATGAAGAACGACAGGTCCGGGGCGAGGCCATTCTCTCAGGACGCGGACCG	725
Qy	1256	AACCTACGACTACGCCAGCAACAGCGTCTTCTGCACCTTGACGCTGGGCGACGAGGTCTTC	1315


```
Db 721 CCGCCAGAGGACAGAGC-ACCTCCGAGGAGATACCCGACCCCAAGTGCCTGGACC 779
1652 CCATAGGGCAGAGTCTGCTCTCTTTTGTACAGAGATGGGAGCAGTTTAAATAG 1711
780 CCATAGGGCAGAGTCTGCTCTCTTTTGTACAGAGATGGGAGCAGTTTAAATAG 839
1712 CGGGACTCAGAGCCAGAAACCGGAGGAGGAGCCCGACCTTCCGAGGGAATAACA 1771
840 CGGGACTCAGAGCCAGAAACCGGAGGAGGAG-CCCGCAGCTTTCGAGGGAATAACA 898
1772 GAACAGGAGGAGCCCAT 1790
899 GAACAGGAGGAGCCCAT 917

RESULT 4
AL529898 942 bp mRNA linear EST 13-FEB-2001
LOCUS AL529898 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD005YH17 5
DEFINITION prime, mRNA sequence.
ACCESSION AL529898
VERSION AL529898.1 GI:12793391
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 942)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..942
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD005YH17"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 160 a 307 g 112 t 6 others
ORIGIN
Query Match 43.1%; Score 868.6; DB 9; Length 942;
Best Local Similarity 98.8%; Pred. No. 4.9e-163;
Matches 890; Conservative 5; Mismatches 4; Indels 2; Gaps 2;

QY 24 GGACCGCGCCGACAGAGAGCGCGGGCCAGGGCTACTAGCAGGACTGGCGCGCGCGC 83
Db 1 GGACCGCGCCGACAGAGAGCGCGGGCCAGGGCTACTAGCAGGACTGGCGCGCGCGC 60
QY 84 AGGGTAGCAAGGTGAGTCGCTTTCGCAAGAGCAGCGCCAAACCTACTAGAGAT 143
Db 61 AGGGTAGCAAGGTGAGTCGCTTTCGCAAGAGCAGCGCCAAACCTACTAGAGAT 120
QY 144 CGCGCCCGGTGAGCAGCACCCGACGTCAGAGCCCGGAGCTCCGAGCGCGGGAGCAG 203
Db 121 CGCGCCCGGTGAGCAGCACCCGACGTCAGAGCCCGGAGCTCCGAGCGCGGGAGCAG 180
```

```
QY 204 TCCCTCTCTCATCAGGGAGTGTCTATCTTGGCAGTCTTGGGACCCAGGACCCGCGGCATC 263
181 TCCCTCTCTCATCAGGGAGTGTCTATCTTGGCAGTCTTGGGACCCAGGACCCGCGGCATC 240
264 CTTGAGAGAGCAGCAGTCTTGGAGAGCAGGCATCTCAGATCCCTTAAGAAACACGCGTCCG 323
241 CTTGAGAGAGCAGCAGTCTTGGAGAGCAGGCATCTCAGATCCCTTAAGAAACACGCGTCCG 300
324 AGAAGCCGCGGATCTCAGGTGCCAGGATCGTTAGGACTGAACGGGAGGGTACTAGAGGA 383
301 AGAAGCCGCGGATCTCAGGTGCCAGGATCGTTAGGACTGAACGGGAGGGTACTAGAGGA 360
384 CCACCTGGCTCTGGACCCGTCGGAGGTGCCCTTGACGTAAACCCAGGAGGGCTCCCTTTG 443
361 CCACCTGGCTCTGGACCCGTCGGAGGTCGCCCTGACGTAAACCCAGGAGGGCTCCCTTTG 420
444 ACGGAGCGCTTGGGAGCGGACCCGCGCTTGGAGCCCGCAGAGCAGGGTAAGGGAG 503
421 ACGGAGCGCTTGGGAGCGGACCCGCGCTTGGAGCCCGCAGAGCAGGGTAAGGGAG 480
504 CGGGGGGACGCGTCCGGGGAGTGCAGACCCAGGCCCCAAGGGGGTCAACCGCTCTCTGGCC 563
481 CGGGGGGACGCGTCCGGGGAGTGCAGACCCAGGCCCCAAGGGGGTCAACCGCTCTCTGGCC 540
564 CGCGAGAGCCCGCGCGCCGCGCATTTGCGCCCAAGAGTGAAGAGATTTGCTGGCCC 623
541 CGCGAGAGCCCGCGCGCCGCGCATTTGCGCCCAAGAGTGAAGAGATTTGCTGGCCC 600
624 TGGCAGCGTCCGGGCTGAGCGCGGCAAGAGGGTGGCGGCGCGCGCTCGGAGTGGCCA 683
601 TGGCAGCGTCCGGGCTGAGCGCGGCAAGAGGGTGGCGGCGCGCGCTCGGAGTGGCCA 660
684 TGGTCTGCTGCTGTGGTGGCCATCCCGCTGCTGTGCACAGCTCCCGCGGCGCCAGCGC 743
661 TGGTCTGCTGCTGTGGTGGCCATCCCGCTGCTGTGCACAGCTCCCGCGGCGCCAGCGC 720
744 ACTACGAGATGCTGGTCTGCTCCGCGATGCTGCGACCCCGATGCGGCGCCCGCTGGCCCTG 803
721 ACTACGAGATGCTGGTCTGCTCCGCGATGCTGCGACCCCGATGCGGCGCCCGCTGGCCCTG 780
804 -GTCCCGCGCGCGCGCTGCTTCCGTCGCGCCCTTCCCGCCAGGCGCCCAAGGAGAGTG 862
781 TGTCCCGCGCGCGCGCTGCTTCCGTCGCGCCCTTCCCGCCAGGCGCCCAAGGAGAGTG 840
863 GGCGCGCGCGGAAAGCAGCGCTTGGCGGGCGCCCTTGACACACAGCTCCCAAGAGGGCCC 922
841 GGCGCGCGCGGAAAGCAGCGCTTGGCGGGG-CCCTTGGCACAACAGTCCCAAGAGGGCCC 899
923 C 923
900 C 900

RESULT 5
AL518440
LOCUS AL518440 LTI_NFL011_NBC1 Homo sapiens cDNA clone CS0DA009YF06 5
DEFINITION prime, mRNA sequence.
ACCESSION AL518440
VERSION AL518440.1 GI:12781933
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 894)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
```



```
RESULT 8
BE888168
LOCUS      601511679P1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3913171 5',
DEFINITION mRNA sequence.
ACCESSION BE888168
VERSION   BE888168.1 GI:10344201
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 387)
           NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC.
           cDNA Library Preparation: Life Technologies, Inc.
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: L1AM9732 row: n column: 20
           High quality sequence stop: 387.
           Location/Qualifiers
             1..387
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:3913171"
               /clone_lib="NIH_MGC_71"
               /tissue_type="leiomyosarcoma"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
               Site_2: SalI; cloned unidirectionally. Primer: Oligo dT.
               Average insert size 2.1 kb."
               /size 106 c 149 g 46 t

BASE COUNT      86 a 106 c 149 g 46 t

FEATURES             source
    source
    Query Match      17.6%; Score 355.8; DB 12; Length 387;
    Best Local Similarity 98.7%; Pred. No. 6.5e-61;
    Matches 380; Conservative 0; Mismatches 2; Indels 3; Gaps 2;

QY 1 AGGAAGTTAGGAGGCGAGAGGACCGCGCCGCA-GAGAGAGGCGCGCGCAGGGCT 59
    |||||||
DB 5 AGGAAGTTAGGAGGCGAGAGGACCGCGCCGCA-GAGAGAGGCGCGCGCAGGGCT- 63
    |||||||

QY 60 ACTAGCAGGGACTGGGCGCGGCGGAGGTAGCAAGTGTGCTGCTTCCCAAGAGGC 119
    |||||||
DB 64 -CTAGCAGGGACTGGGCGCGGCGGAGGTAGCAAGTGTGCTGCTTCCCAAGAGGC 122
    |||||||

QY 120 AGAGCGCAAACTACTAGGAGATCGCGCCCGGTGAGCAGCACCCTGAGCTCAGAGCCCG 179
    |||||||
DB 123 AGAGCGCAAACTACTAGGAGATCGCGCCCGGTGAGCAGCACCCTGAGCTCAGAGCCCG 182
    |||||||

QY 180 GGAGCTCGGAGCGGGGAGCAGTCCCTCTCCATCAGGAGTGGTCTATCTGGGCAGT 239
    |||||||
DB 183 GGAGCTCGGAGCGGGGAGCAGTCCCTCTCCATCAGGAGTGGTCTATCTGGGCAGT 242
    |||||||

QY 240 CTGGGACCCAGGCACCGCGCCATCCCTGAGAGAGCAGCAGTCTGAGAGCAGGATCTCA 299
    |||||||
DB 243 CTGGGACCCAGGCACCGCGCCATCCCTGAGAGAGCAGCAGTCTGAGAGCAGGATCTCA 302
    |||||||

QY 300 GATCCCTAAGAAACAGCCGTCGAGAGAGCCGCGGATCTCAGGTGCCAGGATCTTAGG 359
    |||||||
DB 303 GATCCCTAAGAAACAGCCGTCGAGAGAGCCGCGGATCTCAGGTGCCAGGATCTTAGG 362
    |||||||

QY 360 ACTGACGGGAGGCTACTAGAGGAC 384
    |||||||
DB 363 ACTGACGGGAGGCTACTAGAGGAC 387
    |||||||
```

```
RESULT 9
BM844470
LOCUS      K-EST0122585 S13KMS5 Homo sapiens cDNA clone S13KMS5-36-C10 5',
DEFINITION mRNA sequence.
ACCESSION BM844470
VERSION   BM844470.1 GI:19200869
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 425)
           Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
           Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
           Kim,Y.S.
           21C Frontier Korean EST Project 2001
           Unpublished (2002)
           Contact: Kim YS
           Genome Research Center
           Korea Research Institute of Bioscience & Biotechnology
           52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
           Tel: +82-42-860-4470
           Fax: +82-42-860-4409
           Email: yongsung@mail.kribb.re.kr
           Plate: 36 row: C column: 10
           High quality sequence stop: 425.
           Location/Qualifiers
             1..425
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="S13KMS5-36-C10"
               /clone_lib="S13KMS5"
               /tissue_type="myeloma"
               /cell_line="KMS-5"
               /lab_host="Top10F"
               /note="Vector: pCNS; Site_1: EcoRI; Site_2: NotI; The poly
               (A)+ RNA was dephosphorylated with bacterial alkaline
               phosphatase (BAP) and then decapped with tobacco acid
               pyrophosphatase (TAP). The decapped intact mRNA was
               ligated with DNA-RNA linker including EcoR I site by
               treatment of T4 RNA ligase and the first strand cDNA was
               synthesized from oligo dt-selected mRNA by priming with
               dt-tailed vector. The dt-tailed vector was adjusted to
               have about 60nt. The cDNA vector was circularized with E.
               coli DNA ligase after digestion of EcoRI which site is
               also included in vector. An RNA strand converted to a DNA
               strand by Okayama-Berg method. The obtained cDNA vectors
               were used for transposition of competent cells E. coli
               Top10F' by electroporation method. The cDNA libraries
               constructed by this method are full-length enriched cDNA
               library."
               /size 93 a 119 c 160 g 53 t

BASE COUNT      93 a 119 c 160 g 53 t

ORIGIN
    Query Match      17.5%; Score 353.8; DB 14; Length 425;
    Best Local Similarity 96.2%; Pred. No. 1.6e-60;
    Matches 384; Conservative 0; Mismatches 12; Indels 3; Gaps 2;

QY 1 AGGAAGTTAGGAGGCGGAGAGGACCGCGCCCGCAGAGAGAG--GCGCGGCGCCAGGCG 58
    |||||||
DB 27 AGGAAGTTAGGAGGCGGAGAGGACCGCGCCCGCAGAGAGAGTGCGCGGCGCAGTGT 86
    |||||||

QY 59 TACTAGCAGGACCTGGGGCGCGGCGGAGGGGTAGCAAGTGTAGTGGTCTTGCCTCAAGAGG 118
    |||||||
DB 87 CTCTAGCAGGACCTGGGGCGCGGCGGAGGGGTAGCAACGTGAGTGGTCTTGCCTCAAGAGG 146
    |||||||

QY 119 CAGAGCGCAAACTACTAGGAGATCGCGCCCGGTGAGCAGCACCCTGAGCTCAGAGCCC 178
    |||||||
DB 147 CAGAGCGCAAACTACTAGGAGATCGCGCCCGGTGAGCAGCACCCTGAGCTCAGAGCCC 206
    |||||||
```


plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 110 a 209 c 145 g 88 t

ORIGIN

Query Match 15.6%; Score 314.4; DB 13; Length 552;
Best Local Similarity 76.8%; Pred No. 1.2e-52;
Matches 394; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 1001 GTGCTCGCATGCTTTCTACCGGCGCTCGCGGCGCCACGAGGGTTACGAGTGCTG 1060
DB 8 GTGCGCGGCTGCTTCTACGCGGCTCTCAAGAACCCACGAGGGTTACGAGTACTC 67
QY 1061 CGCTTCGAGCGTGGTACCAACCTGGCAACGCTACGAGGACGACCGGCAAGTTT 1120
DB 68 AAGTTTGAGCGTGGTACCAACCTAGCAACACTAGCGCGGCGGCGGCAAGTTT 127
QY 1121 ACTTGCCCATGCCAGGCGTCTACTTCTTCGCTTACCAGTGTCTCATGCGGCGGCGAC 1180
DB 128 AGTGCACATTCGCGGACCTACTTTTTCACCTACCATGTCTCTATGCGGCGGCGAC 187
QY 1181 GGCACAGCATGTGGCGGACCTCATGAAGACGACAGGTCCGGGCGGCGGCGGCGG 1240
DB 188 GGCACAGCATGTGGCGGACCTCTGCAAGATGCGGCGGCGGCGGCGGCGGCGGCGG 247
QY 1241 CAGGACGCGGACGACACTACGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1300
DB 248 CAGGACGCGGACGACACTACGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 307
QY 1301 GCGACGAGGCTCTCATCAAGCTGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1360
DB 308 GCGACGAGGCTCTCATCAAGCTGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 367
QY 1361 TACAGACCTTCTCGGCTTATCATCTACCCGCTAGCGGCGGCGGCGGCGGCGGCGGCGG 1420
DB 368 TACAGACCTTCTCGGCTTATCATCTACTCCGCTAGCTCCGCGGCGGCGGCGGCGGCGG 427
QY 1421 CGCTGCGGCTTCTCTCGGCTCTCAGCCACCTCTGCGGCGGCGGCGGCGGCGGCGGCGG 1480
DB 428 CCAGTGCCTTACCGCGGCTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 487
QY 1481 CCACCTTTTGAAGCGCTGGC 1500
DB 488 CCACCGGCTCGCTGCGGCGG 507

RESULT 13

AQ200834/c

LOCUS

DEFINITION AQ200834 323 bp DNA linear GSS 20-APR-1999
RPC111-62K10.TJ RPC1-11 Homo sapiens genomic clone RPC1-11-62K10,
DNA sequence.

ACCESSION

AQ200834

VERSION

AQ200834.1

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,

Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.

Use of human BAC End Sequences for Sequence-Ready Map Building

Unpublished (1998)

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

FEATURES

source

1. 323
/organism="Homo sapiens"
/db_xref="GDB:7523673"
/db_xref="taxon:9606"
/clone="RPC1-11-62K10"
/clone_lib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT 61 a 91 c 87 g 84 t

ORIGIN

Query Match 15.5%; Score 313.4; DB 17; Length 323;
Best Local Similarity 98.1%; Pred No. 1.8e-52;
Matches 317; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1605 AGAGCGACTTCGGAGGATCACCCGACCCCAAGTGCCTGAGCCCATAGGGGCGAGA 1664
DB 323 AGAGCGACTTCAGAGGGATCACCCGACCCCAAGTGCCTGAGCCCATAGGGGCGAGA 264
QY 1665 GGTCTGGCTTCTCTTTGTACAGATGGGGAGCAGTTTAAATAGCGGGACTCAGAGG 1724
DB 263 GGTAGTGGCTTCTCTTTGTACAGATGGGGAGCAGTTTAAATAGCGGGACTCAGAGG 204
QY 1725 CCCAGAAAGCGGAGGAGGAGCCCGCAGCTTTCGAGGGAATAACAGAAACAGGAGGAG 1784
DB 203 CCCAGAAAGCGGAGGAGGAGCCCGCAGCTTTCGAGGGAATAACAGAAACAGGAGGAG 144
QY 1785 CCCATTTAGGCAAGAGAAGACATTAACAGAGGTAGTGCAGGTTCCTCCTCACAATTC 1844
DB 143 CCCATTTAGGCAAGAGAAGACATTAACAGAGGTAGTGCAGGTTCCTCCTCACAATTC 84
QY 1845 TCTCCGCCCTCTCGTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCC 1904
DB 83 TCTCCGCCCTCTCGTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCC 24
QY 1905 CTCTCTGTGAAGTGGAGGAACCA 1927
DB 23 CTCTCTGTGAAGTGGAGGAACCA 1

RESULT 14

BI793031

LOCUS

DEFINITION

BI793031

ACCESSION

BI793031

VERSION

BI793031.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

1 (bases 1 to 461)

Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,

Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,

Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,

Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas

, M., Gibbons, M., McCann, R., Cole, R., Tsagarelshvili, R., Williams, T.

, Jackson, Y. and Bowers, Y.

Endocrine Pancreas Consortium

Unpublished (2000)

Other_ESTs: ie47a04.xl

QY	1421	CGCTCGCCCTTCTCTCCCG	1440	
Db	428	CCACGTCCTCTCACCCGCCGG	447	
RESULT	15			
LOCUS	BM354017			
DEFINITION	BM354017 540 bp mRNA linear EST 12-MAR-2002			
	if29d01.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens			
	cDNA clone IMAGE:5677849 5' similar to TR:075973 075973 C1Q-RELATED			
	FACTOR. ;, mRNA sequence.			
ACCESSION	BM354017			
VERSION	BM354017.1	GI:18086684		
KEYWORDS	EST.			

ESI.
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 540)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Secaere, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
, M., Gibbons, M., McCann, R., Cole, R., Tsagarieishvili, R., Williams, T.,
Jackson, Y. and Powers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Other_ESTS: if29d01.xl
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohpc.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:

```

Washington University Genome Sequencing Center for information on
obtaining a clone please contact: Juliana Brown
(brownjefas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 451.

FEATURES             Location/Qualifiers
     1..540
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="IMAGE:5677849"
     /clone_lib="Wellton Normalized Human Islet 4 N4-HIS 1"
     /sex="Both"
     /tissue_type="Islets of Langerhans"
     /dev_stage="Adult"
     /lab_host="DH10B"
     /notes="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
     Site_2: Sal 1; Starting library constructed using
     SuperScript Plasmid Library kit (Life Technologies). cDNA
     made by oligo-dT priming. Size-selected by column
     fractionation; average insert size 1.08 kb. Library was
     amplified once on solid support and plasmid DNA from
     library was prepared. The library DNA was normalized by
     method #4 from Bonaldo, Lennon, and Soares 1996 Genome
     Research 6:791-806; 0.5 microgram single-stranded library
     plasmid DNA was mixed with 5 microgram PCR product
     representing library inserts and hybridized to an Ecot of
     20. Single-stranded (unhybridized) plasmids were isolated
     by hydroxyapatite chromatography and used to make this
     library."

BASE COUNT      108 a      203 c      142 g      87 t

Query Match          15.3%   Score 307.6;  DB 13;   Length 540;
Best Local Similarity 76.7%   Pred. No. 2.6e-51;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 16:36:13 ; Search time 72 Seconds
(without alignments)
440.467 Million cell updates/sec

Title: US-10-005-499-378
Perfect score: 1311
Sequence: 1 MVLALLVAIPLLVHSSRGA.....VHGNTNKYSTFGFIIPD 238

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1311	100.0	238	23	AAU84377
2	1286.5	98.1	267	23	AAU84378
3	907.5	69.2	255	22	AAU84379
4	903.5	68.9	255	23	AAU84380
5	869.5	66.3	287	23	AAU84381
6	662.5	50.5	225	23	AAU84382
7	376	28.7	744	22	AAU84383
8	376	28.7	755	22	AAU84384
9	352	26.8	253	22	AAU84385
10	352	26.8	255	22	AAU84386

11	345.5	26.4	638	23	ABJ01025	Human breast speci
12	345	26.3	635	22	AAU84377	Human protein SEQ
13	345	26.3	644	22	AAU84378	Human protein SEQ
14	329.5	25.1	314	23	ABP41487	Human ovarian anti
15	329	25.1	285	21	AAU84377	Murine ACRP30R1M.
16	328	25.0	285	21	AAU84377	Human adipocyte co
17	328	25.0	285	21	AAU84377	A polypeptide desi
18	328	25.0	285	21	AAU84377	Human adipocyte co
19	327	24.9	245	21	AAU84377	Human OREF ORF2827
20	327	24.9	245	21	AAU84377	Human polypeptide
21	327	24.9	277	22	AAU84377	Human polypeptide
22	327	24.9	285	21	AAU84377	Rat skin cell secr
23	327	24.9	285	22	AAU84377	Skin cell protein,
24	327	24.9	285	23	ABU72242	Rat protein isolat
25	327	24.9	294	21	AAU84377	Rat skin cell prot
26	327	24.9	294	22	AAU84377	Skin cell protein,
27	327	24.9	294	23	ABU72178	Rat protein isolat
28	327	24.9	331	22	AAU84377	Human diagnostic a
29	324.5	24.8	244	21	AAU84377	Human adiponectin
30	322	24.6	244	18	AAU84377	Human complement c
31	322	24.6	245	21	AAU84377	Human Cql C protei
32	322	24.6	245	22	AAU84377	Human Cql C protei
33	322	24.6	245	22	AAU84377	Human Cql C protei
34	320	24.4	244	21	AAU84377	Human APM1 (Adipos
35	319.5	24.4	289	22	AAU84377	Mouse ZACRP7. Mus
36	319.5	24.4	289	23	ABU72306	Adipose most abund
37	318	24.3	244	20	AAU84377	Rat protein isolat
38	318	24.3	244	21	AAU84377	Human adipocyte co
39	318	24.3	244	21	AAU84377	Human APM1 (Adipos
40	318	24.3	244	22	AAU84377	Human OBG3 protein
41	318	24.3	244	22	AAU84377	Human adipocyte co
42	318	24.3	244	22	AAU84377	Human ACRP30 prote
43	318	24.3	244	22	AAU84377	Human adipocyte co
44	318	24.3	244	23	ABU82223	Human apml protein
45	318	24.3	244	23	AAU84377	Human APM1 protein

ALIGNMENTS

RESULT 1
AAU84377
ID AAU84377 standard; Protein; 238 AA.
XX
AC AAU84377;
XX
DT 08-MAY-2002 (first entry)
XX
DE Novel human secreted or membrane-associated protein #16.

Human; secreted protein; membrane-associated protein; hypertension;
inflammatory disorder; neurological disorder; hematopoietic disorder;
skeletal developmental disorder; growth abnormality; autoimmune disorder;
neurodegenerative disorder; nervous system disorder; bacterial infection;
peripheral myelinopathy; viral infection; cancer; obesity; diabetes;
hypotension; sexual development disorder; blood disorder.

OS Homo sapiens.

XX WO200204600-A2.

XX 17-JAN-2002.

XX 12-JUL-2001; 2001WO-US21985.

XX 12-JUL-2000; 2000US-218033P.

XX 21-AUG-2000; 2000US-226517P.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (GLAX) SMITHKLINE BEECHAM PLC.

XX AGARWAL P, Cogswell JP, Lai Y, Martensen SA, Rizvi SK, Strum JC;

```
PI Smith RF, Xiang Z, Xie Q;
XX WPI; 2002-188468/24.
DR N-PSDB; ABK35597.
XX
XX Novel secreted and membrane-associated polypeptides and polynucleotides
PT encoding the polypeptides, for preventing, treating and ameliorating
PT cancers, mental or sexual developmental disorders, and malignant tumours
PT
XX
XX Claim 1; Page 139-140; 151pp; English.
PS
XX The present invention relates to the isolation of novel human secreted
XX or membrane-associated proteins and the genes encoding them. The
XX sequences of the invention are useful for treating, preventing and
XX ameliorating various diseases such as inflammatory disorders (e.g.
XX asthma), neurological disorders (e.g. dementia), haematopoietic
XX disorders, skeletal developmental disorders, growth abnormalities,
XX neurodegenerative disorders (e.g. Huntington's disease), nervous system
XX disorders, autoimmune disorders (e.g. rheumatoid arthritis),
XX peripheral myelinopathies, viral and bacterial infections,
XX alpha-mannosidosis, diabetes, cancers, malignant tumours, hyper- and
XX hypotension, obesity, bulimia, anorexia, manic depression, delirium,
XX mental retardation, Tourette's syndrome, schizophrenia, growth, mental
XX or sexual development disorders, and dysfunctions of the blood cascade
XX system including those leading to stroke. AAU84362-AAU84389 represent
XX the novel human secreted or membrane-associated proteins of the
XX invention.
XX
XX Sequence 238 AA;
XX
XX Query Match 100.0%; Score 1311; DB 23; Length 238;
XX Best Local Similarity 100.0%; Pred. No. 8.3e-101;
XX Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MVLILLVATPLLVHSSRGPAHYEMLGRCRMVCDPHGPGPDGAPASVPPPPGAKGEV 60
XX DB 1 MVLILLVATPLLVHSSRGPAHYEMLGRCRMVCDPHGPGPDGAPASVPPPPGAKGEV 60
XX
XX QY 61 GRRGKAGLRGPPGPGPRGPGPGPPGPGGVAAGVVPRIAFYAGLRPH 120
XX DB 61 GRRGKAGLRGPPGPGPRGPGPGPPGPGGVAAGVVPRIAFYAGLRPH 120
XX
XX QY 121 EGYEVLRFDDVTNVNGNAYEAASGKFTCPMPGVYFFAYHVMRGDGTSMWADLMKNGOV 180
XX DB 121 EGYEVLRFDDVTNVNGNAYEAASGKFTCPMPGVYFFAYHVMRGDGTSMWADLMKNGOV 180
XX
XX QY 181 RASATQAQDADQNYDVASNSVILHLDVGDEVFIKLDGKRVHGGNTNKYSTFGFIYPD 238
XX DB 181 RASATQAQDADQNYDVASNSVILHLDVGDEVFIKLDGKRVHGGNTNKYSTFGFIYPD 238
XX
XX RESULT 2
XX AAU84378
XX ID AAU84378 standard; Protein; 267 AA.
XX
XX AC AAU84378;
XX
XX DT 08-MAY-2002 (first entry)
XX
XX DE Novel human secreted or membrane-associated protein #17.
XX
XX KW Human; secreted protein; membrane-associated protein; hypertension;
XX inflammatory disorder; neurological disorder; haematopoietic disorder;
XX skeletal developmental disorder; growth abnormality; autoimmune disorder;
XX neurodegenerative disorder; nervous system disorder; bacterial infection;
XX peripheral myelinopathy; viral infection; cancer; obesity; diabetes;
XX hypotension; sexual development disorder; blood disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200204600-A2.
XX
XX AC
```

```
PD 17-JAN-2002.
XX
XX 12-JUL-2001; 2001WO-US21985.
XX
XX 12-JUL-2000; 2000US-218033P.
XX
XX 21-AUG-2000; 2000US-226517P.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX (SMIK ) SMITHKLINE BEECHAM PLC.
XX (GLAX ) GLAXO GROUP LTD.
XX
XX Agarwal P, Cogswell JP, Lai Y, Martensen SA, Rizvi SK, Strum JC;
XX Smith RF, Xiang Z, Xie Q;
XX
XX WPI; 2002-188468/24.
XX N-PSDB; ABK35598.
XX
XX Novel secreted and membrane-associated polypeptides and polynucleotides
XX encoding the polypeptides, for preventing, treating and ameliorating
XX cancers, mental or sexual developmental disorders, and malignant tumours
XX
XX Claim 1; Page 140; 151pp; English.
XX
XX The present invention relates to the isolation of novel human secreted
XX or membrane-associated proteins and the genes encoding them. The
XX sequences of the invention are useful for treating, preventing and
XX ameliorating various diseases such as inflammatory disorders (e.g.
XX asthma), neurological disorders (e.g. dementia), haematopoietic
XX disorders, skeletal developmental disorders, growth abnormalities,
XX neurodegenerative disorders (e.g. Huntington's disease), nervous system
XX disorders, autoimmune disorders (e.g. rheumatoid arthritis),
XX peripheral myelinopathies, viral and bacterial infections,
XX alpha-mannosidosis, diabetes, cancers, malignant tumours, hyper- and
XX hypotension, obesity, bulimia, anorexia, manic depression, delirium,
XX mental retardation, Tourette's syndrome, schizophrenia, growth, mental
XX or sexual development disorders, and dysfunctions of the blood cascade
XX system including those leading to stroke. AAU84362-AAU84389 represent
XX the novel human secreted or membrane-associated proteins of the
XX invention.
XX
XX Sequence 267 AA;
XX
XX Query Match 98.1%; Score 1286.5; DB 23; Length 267;
XX Best Local Similarity 89.1%; Pred. No. 1e-98;
XX Matches 238; Conservative 0; Mismatches 0; Indels 29; Gaps 1;
XX
XX QY 1 MVLILLVATPLLVHSSRGPAHYEMLGRCRMVCDPHGPGPDGAPASVPPPPGAKGEV 60
XX DB 1 MVLILLVATPLLVHSSRGPAHYEMLGRCRMVCDPHGPGPDGAPASVPPPPGAKGEV 60
XX
XX QY 61 GRRGKAGLRGPPGPGPRGPGPGPPGPGGVAAGVVPRIAFYAGLRPH 120
XX DB 61 GRRGKAGLRGPPGPGPRGPGPGPPGPGGVAAGVVPRIAFYAGLRPH 120
XX
XX QY 92 PFGPGGGVAPAAAGVVPRIAFYAGLRPHGEGVEVLRFDVTVNGNAYEAASGKFTCPMP 151
XX DB 121 PFGPGGGVAPAAAGVVPRIAFYAGLRPHGEGVEVLRFDVTVNGNAYEAASGKFTCPMP 180
XX
XX QY 152 GYVFFAYHVMRGDGTSMWADLMKNGOVRAAIAQADADQNYDVASNSVILHLDVGDEVF 211
XX DB 181 GYVFFAYHVMRGDGTSMWADLMKNGOVRAAIAQADADQNYDVASNSVILHLDVGDEVF 240
XX
XX QY 212 IKLDGKRVHGGNTNKYSTFGFIYPD 238
XX DB 241 IKLDGKRVHGGNTNKYSTFGFIYPD 267
XX
XX RESULT 3
XX AAG64212.
XX ID AAG64212 standard; Protein; 255 AA.
XX
XX AC AAG64212;
```

AC: AAG64212

19-SEP-2001 (first entry)
MURINE HSP47 INTERACTING PROTEIN, #2.
MURINE; HEAT SHOCK PROTEIN INTERACTING PROTEIN; HSP47.

Mus sp.

JP2001145493-A.

29-MAY-2001.

19-NOV-1999; 99JP-0330631.

19-NOV-1999; 99JP-0330631.

(KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

WPI; 2001-395263/42.

N-PSDB; AAH48068.

Using the two-hybrid screening method to prepare proteins which interact with the heat shock protein HSP47.

Claim 7; Page 23; 26pp; Japanese.

The present invention relates to a method for preparing a protein which interacts with the heat shock protein HSP47. The method involves the two-hybrid screening method using the HSP47 gene and a mammalian cDNA library. The present sequence is the protein sequence for a murine HSP47 interacting protein, which was used in the present method. The HSP47 interacting proteins are useful for the diagnosis and treatment of diseases caused by an increase or decrease in activity of HSP47.

Sequence 255 AA;

Query Match 69.2%; Score 907.5; DB 22; Length 255;
Best Local Similarity 70.2%; Pred. No. 2.2e-67;
Matches 179; Conservative 18; Mismatches 41; Indels 17; Gaps 7;
QY 1 MVLILLVAIPLLVHSSRGPAHYEMLCRCMVCDPH-GPRGPG-PPGAPSVPPPPP 54
DB 1 MVLILLVILPLVSSAGTSAHYEMLCRCMVCDPYGCTKAPSTAATPDRLGMSLPTFIQ 60
QY 55 GAKGEVGRGKAGLRGP---PGPPGPGPP---GEPGR---PGPPGPGPGGVAPAG 105
DB 61 GPKGGRGPGKAGRGPCEGPGPGPGPGPGKEGPGQGLPGPPGAPGLNAAGISAAT 120
QY 106 Y--VPRIAFYAGLRPHGEYEVLRFDVVTVNGVAYEASGKFTCPMPGVYFFAYHVLNR 163
DB 121 YSTVPKIAFYAGLRQHEGYEVLFKFDVVTVNLGNHYDPTTKFTCSIPGIYFFTVHVLNR 180
QY 164 GGDGTSWADLMKNGQVRASIAQADADQNYDASNVILHLDVGEVEIKLDGKVGCGN 223
DB 181 GGDGTSWADLMKNGQVRASIAQADADQNYDASNVILHLDVGEVEIKLDGKVGCGN 240
QY 224 TNKYSTFGFIYIPD 238
DB 241 NNKYSTFGFIYAD 255

RESULT 4

ABB53290

ID ABB53290 standard; Protein; 255 AA.

XX ABB53290;

DF 12-FEB-2002 (first entry)

DE Human polypeptide #30.

XX Human; nootropic; neuroprotective; anticonvulsant; antidepressant;

neuroleptic; tranquiliser; antiarrhythmic; cardiant; antiasthmatic;
antiinflammatory; antilipaeic; hepatotropic; virucide; antidiabetic;
nephrotropic; anorectic; cytostatic; vaccine; neurologic disease;
cardiovascular disease; respiratory disease; liver disease;
renal disease; skeletal muscle disease; gastrointestinal disease;
placental disease; testicular cancer; male fertility; pancreatic disease.

Homo sapiens.

WO200181363-A1.

01-NOV-2001.

26-APR-2001; 2001WO-US13360.

27-APR-2000; 2000US-199963P.

11-MAY-2000; 2000US-20336P.

25-MAY-2000; 2000US-207087P.

26-MAY-2000; 2000US-207546P.

(SMIK) SMITHKLINE BEECHAM CORP.
(SMIK) SMITHKLINE BEECHAM PLC.

Agarwal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;
Lai Y, Xie Q;

WPI; 2002-041392/05.

N-PSDB; ABA90355.

Novel polypeptides and polynucleotides useful as a vaccine for preventing and treating diseases associated with the polypeptide, e.g. Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility, asthma, amnesia.

Claim 1; Page 101; 116pp; English.

The invention relates to an isolated polypeptide comprising a 277, 480, 583, 581, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026, 844, 782, 628, 394, 471, 485, 286, 533, 495, 350, 619, 490, 462, 255, 784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as given in the specification. The polypeptides, modulators of the polypeptides and antibodies against the polypeptides are useful for treating diseases such as neurological and psychiatric diseases including Alzheimer's, paraneoplastic, Huntington's disease, myotonic dystrophy, anorexia and depression; cardiovascular diseases including congestive heart failure, Hodgkin's disease and myocardial infarction; respiratory diseases including asthma, chronic obstructive pulmonary disease, cystic fibrosis and adult respiratory distress syndrome; liver diseases including hypercholesterolaemia, cirrhosis, viral and nonviral hepatitis, Type II diabetes mellitus, and impaired glucose tolerance; renal disease including renal failure, acute tubular necrosis and glomerulonephritis; skeletal muscle diseases including Eulenburg's disease, hypoglycaemia and obesity; gastrointestinal diseases including myotonia congenita and intestinal obstruction; lymph diseases including lymphagiectasia; diseases of placenta including chorioncarcinoma; diseases of testes including testicular cancer, male reproductive diseases including low testosterone and male infertility; and disease of pancreas including diabetic ketoacidosis, Type 1 and 2 diabetes and obesity. The present sequence is a polypeptide of the invention.

Sequence 255 AA;

Query Match 68.9%; Score 903.5; DB 23; Length 255;
Best Local Similarity 69.8%; Pred. No. 4.8e-67;
Matches 178; Conservative 18; Mismatches 42; Indels 17; Gaps 7;
QY 1 MVLILLVAIPLLVHSSRGPAHYEMLCRCMVCDPH-GPRGPG-PPGAPSVPPPPP 54
DB 1 MVLILLVILPLVSSAGTSAHYEMLCRCMVCDPYGCTKAPSTAATPDRLGMSLPTFIQ 60
QY 55 GAKGEVGRGKAGLRGP---PGPPGPGPP---GEPGR---PGPPGPGPGGVAPAG 105

The invention relates to an isolated novel secreted polypeptide (I) and polynucleotide (II). (I) and (II) are useful for treating cancer, autoimmune diseases, wound healing disorder, infections, haematopoietic disorders, inflammatory disorders, infertility, neurological and psychiatric diseases, cardiovascular diseases, respiratory diseases, renal diseases, or gastrointestinal diseases. These may also be used to treat diseases, abnormalities and disorders caused by abnormal expression, production, function and/or metabolism of the genes, as vaccines for inducing immunological response in a mammal, and in screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. The polypeptides can be used as immunogens to produce antibodies immunospecific for the polypeptides, and to identify membrane-bound or soluble receptors. The polynucleotides may be used as diagnostic reagents, in chromosome localisation studies, and in tissue expression studies. The present sequence represents the amino acid sequence of novel human secreted protein #6.

The invention relates to a human C1q-related factor-like (CRF-like) protein, LP231, which has sequence homology to human cerellin, and its associated nucleic acid. The polypeptides and polynucleotides are useful for treating or preventing neurological disorders associated with the inappropriate expression of cerellin-2 proteins and disruption of the

DR N-PSDB; AAI58283.

FD XX
1007-7007

CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
XX	
SQ	Sequence 255 AA; 26.8%; Score 352; DB 22; Length 255;
	Query Match 35.9%; Pred. No. 2.1e-21;
	Best Local Similarity 34; Mismatches 90; Indels 42; Gaps 10;
	Matches 93; Conservative

[illegible]

Db 66 LPGLAGDHGEFGEKGDPIPGNPGKVGPKGPMGPKGGPGAPGAPGKGESGDYKA--TQ 122

OV 109 RIAFYA-----GLRRPHEGYEVLRFDDVVTNVGNAYEAASGKFTCPMPGVYFFAYHVLM 162

db 123 KIAFSATRTINVPLRDR-----OTIRFDHVITNMNNNYEPRSGKFTCKVPGLYFTYHASS 178

163 BGGDGTSMWADI MKNGOVRASATAONADONVND--YASNSVTIHL.DVGDFEVEIKI.DGGKV 219

Db 179 RG---NLCVNLNR-GRERAQKVTFCDYAYNTFQTGGWVLKLGQGVNFQAATDKNS 233
 || :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 220 HGGTNNKYSTFGFIYPD 238
 | |||...||

DB 234 LLGMEGANSTFSGFLFLFPD 232

RESULT 11

ABJ01025
ID ABJ01025 standard. Protein: 638 AA

XX
AC ABJ01025:XX
DT 05-SEP-2002 (first entry)

XX
DE Human breast specific protein SEO ID NO: 108.

XX

KW Human: breast specific gene: breast cancer: gene therapy: breast disease:

KW
XX
cystostatic.

US Homo sapiens.
XX

PN
XX
W02002406/Z-AZ.

FD 23-MAY-2002.
XX
XX
PF 20-NOV-2001; 2001WO-US45079.

XX
PR 20-NOV-2000; 2000US-249998P.

FR
XX
ZZ-NOV-2000; 200003-232503F.

FA (DIAD-) DIADSAUS INC.
XX

PI
XX

DK
XX
LF1; 2002-300230/33.
PT Novel breast-specific polypeptides and polynucleotides encoding

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 16:36:54 ; Search time 21 seconds
(without alignments)
470.065 Million cell updates/sec

Title: US-10-005-499-378

Perfect score: 1311

Sequence: 1 MVLALLVAIPLLVHSSRGPA.....VHGGNTNKYSTFGFIYPD 238

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	918	70.0	258	1	CLRF_MOUSE
2	917	69.9	258	1	CLRF_HUMAN
3	907.5	69.2	255	1	GLIC_MOUSE
4	397	30.3	743	1	CA18_MOUSE
5	388.5	29.6	744	1	CA18_RABIT
6	376	28.7	680	1	CA1A_MOUSE
7	376	28.7	744	1	CA18_HUMAN
8	362.5	27.7	680	1	CA1A_HUMAN
9	360.5	27.5	674	1	CA1A_CHICK
10	360	27.5	674	1	CA1A_BOVIN
11	352	26.8	251	1	CLQB_HUMAN
12	345	26.3	635	1	CA28_HUMAN
13	328	25.0	285	1	COT2_HUMAN
14	327	24.9	245	1	CLQC_HUMAN
15	318	24.3	244	1	APML_HUMAN
16	317.5	24.2	247	1	APML_MOUSE
17	316.5	24.1	289	1	COT7_HUMAN
18	316	24.1	246	1	CLQC_MOUSE
19	316	24.1	253	1	CLQB_RAT
20	314	24.0	245	1	CLQA_HUMAN
21	302.5	23.1	253	1	CLQB_MOUSE
22	302	23.0	245	1	CLQA_MOUSE
23	300	22.9	243	1	COT5_HUMAN
24	287	21.9	246	1	COT3_HUMAN
25	285.5	21.8	419	1	COLE_LEPMA
26	288.5	20.5	215	1	HP25_TAMSI
27	261.5	19.9	215	1	HP27_TAMSI
28	243	18.5	224	1	CERL_RAT
29	239	18.2	278	1	COT6_HUMAN
30	231	17.6	193	1	CERB_MOUSE
31	230.5	17.6	201	1	CERL_HUMAN
32	229	17.5	193	1	CERB_HUMAN
33	222.5	17.0	196	1	HP20_TAMSI

34	216.5	16.5	281	1	COT1_HUMAN
35	214.5	16.4	1453	1	CA11_CHICK
36	211.5	16.1	1464	1	CA11_HUMAN
37	210.5	16.1	1460	1	CA11_CANFA
38	209.5	16.0	675	1	CA39_CHICK
39	208	15.9	453	1	CA44_BOVIN
40	208	15.9	1466	1	CA13_HUMAN
41	207	15.8	1527	1	CA1H_MOUSE
42	206.5	15.8	1459	1	CA12_MOUSE
43	206.5	15.8	1516	1	CA1H_HUMAN
44	206	15.7	329	1	COT4_HUMAN
45	205.5	15.7	779	1	CA11_BOVIN

ALIGNMENTS

RESULT 1
CLRF_MOUSE
ID CLRF_MOUSE STANDARD; PRT; 258 AA;
AC O88992;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
-DE Clq-related factor precursor.
GN CLQRF OR CRF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99097006; PubMed=9878755;
RA Berube N.G., Swanson X.H., Bertram M.J., Kittle J.D., Didenko V.,
RA Baskin D.S., Smith J.R., Pereira-Smith O.M.;
RT "Cloning and characterization of CRF, a novel Clq-related factor,
RT expressed in areas of the brain involved in motor function.";
RL Brain Res. Mol. Brain Res. 63:233-240(1999).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAINSTEM. MORE ABUNDANT IN AREAS
CC OF THE NERVOUS SYSTEM INVOLVED IN MOTOR FUNCTION, SUCH AS THE
CC PURKINJE CELLS OF THE CEREBELLUM, THE ACCESSORY OLIVARY NUCLEUS,
CC THE PONS AND THE RED NUCLEUS.
CC -!- SIMILARITY: CONTAINS 1 COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF095155; AAC64187.1; -.
CC MGD; MGI:1344400; Clqrf.
CC InterPro; IPR001073; Clq.
CC InterPro; IPR000087; Collagen.
CC Pfam; PF00386; Clq; 1.
CC Pfam; PF01391; Collagen; 1.
CC PRINTS; PR00007; COMPLEMENTC1Q.
CC SMART; SM01110; C1Q; 1.
CC PROSITE; PS01113; C1Q; 1.
CC Collagen; Signal.
CC SIGNAL 1 16 POTENTIAL.
CC CHAIN 17 258 C1Q-RELATED FACTOR.
CC DOMAIN 67 115 COLLAGEN-LIKE.
CC DOMAIN 123 258 C1Q.
CC SEQUENCE 258 AA; 26485 MW; F776E2D206EBF763 CRC64;
Query Match 70.0%; Score 918; DB 1; Length 258;
Best Local Similarity 68.6%; Pred. No. 4.8e-53;
Matches 177; Conservative 16; Mismatches 45; Indels 20; Gaps 4;

Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 Hayshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 Hayashizaki Y.
 "Functional annotation of a full-length mouse cDNA collection.";
 Nature 409:685-690(2001).
 -!- FUNCTION: MAJOR COMPONENT OF THE DESCEMET'S MEMBRANE (BASEMENT
 MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
 -!- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
 WITH ALPHA 2(VIII) TYPE COLLAGENS.
 -!- TISSUE SPECIFICITY: HIGH LEVELS IN CALVARIUM, EYE & SKIN OF
 NEWBORN MICE; ALSO IN VARIOUS EPITHELIAL, ENDOTHELIAL AND
 MESENCHYMAL CELLS.
 -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 -!- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.
 -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

 EMBL; X66976; CAA47387.1;
 EMBL; X66977; CAA47387.1; JOINED.
 EMBL; AK018742; BAB31383.1;
 PIR; S23779; S23779.

DR IntelPio; IPR001073; Clq.
DR Interpro: IPR000087: Collagen

```

DR      interflor; AF000008; Collagen.
DR      Pfam: PF00386; C1q; 1.
DR      Pfam: PF01391; Collagen; 7.
DR      PRINTS; PR00007; COMPLEMENTC1Q.
DR      SMART; SM00110; C1q; 1.
DR      PROSITE; PS01113; C1q; 1.
KW      Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
FT      Glycoprotein; Cell adhesion; Collagen; Signal.
KW      SIGNAL      1      28
FT      CHAIN      29      743      COLLAGEN ALPHA 1(VIII) CHAIN.
FT      DOMAIN      29      117      NONHELICAL REGION (NC2).
FT      DOMAIN      118      571      TRIPLE-HELICAL REGION (COL1).
FT      DOMAIN      572      743      NONHELICAL REGION (NC1).
FT      DOMAIN      608      743      C1Q.
FT      CONFLICT      6      6      G -> R (IN REF. 2).
FT      CONFLICT      85      85      H -> Y (IN REF. 1).
FT      CONFLICT      109      109      K -> KG (IN REF. 2).
FT      CONFLICT      248      248      P -> L (IN REF. 1).
FT      CONFLICT      313      313      P -> A (IN REF. 2).
FT      CONFLICT      323      324      IP -> SR (IN REF. 1).
FT      CONFLICT      361      361      D -> H (IN REF. 1).
FT      CONFLICT      596      596      T -> P (IN REF. 2).
FT      CONFLICT      717      719      MPS -> NPF (IN REF. 1).
SQ      SEQUENCE      743 AA; 73454 MW; F584D85BD53897F4 CRC64;

Query Match      30.3%; Score 397; DB 1: Length 743;
Best Local Similarity      38.1%; Pred. No. 6.2e-19;
Matches      94; Conservative.      31; Mismatches      76; Indels      46; Gaps      10;

QY      31      VCDPHGPRG-----PGPDGAPASVPPPPGAKGVRRGKAGLRGPPGPPGPPGPPGPR      86
           : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      502      IGVGSGPIGPGIPGPKGPEGL--PGPPGPPG--VGKPGVAGLHGPCKGALGPGQPGQL      558

QY      87      PGPPGPPG--PGPGGV-----APAGYVPRIAF-----YAGLRPHEG--YEV-----      135
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      559      PGPPGPPGPPGPPAVMPTSPQGEYLPLDMGLGDGVKTPHAYAGKKGKGGPAYEMPAPT      618

QY      126      -----LRDDVVTVNGVAYEAASGKFTCPMPCVYFFAYHVMRGDGGTSMWA      172
           : : : : : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      619      AELTVPPPPVGAPVKFKDKLLYNGQNYPNTGFTCEVPGVYIFAYHVHCKGG----NVWV      675

QY      173      DLWMKNGVRSAAIAQADQNDYASNSVILHLHDVDGDEVFKLDGGKVGHGNTNKY--STF      230

```



```
CC CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC CC -1- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIIII COLLAGENS.
CC CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensed@isb-sib.ch).
CC -----
DR EMBL; X67348; CAA47763.1; --
DR EMBL; X65121; CAA46237.1; --
DR EMBL; X63013; CAA44741.1; --
DR EMBL; Z21610; CAA79736.1; --
DR PIR; S28807; S28807.
DR PIR; S31216; S31216.
DR PIR; S22215; S22215.
DR MGI; MG1:88445; Coll10al.
DR InterPro; IPR001073; Clq.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; Clq; 1.
DR Pfam; PF01391; Collagen; 9.
DR PRINTS; PR00007; COMPLENNTC1Q.
DR ProDom; PD000007; Collagen; 2.
DR SMART; SM00110; Clq; 1.
DR PROSITE; PS01113; C1Q; 1.
DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cartilage; Collagen; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 680 COLLAGEN ALPHA 1(X) CHAIN.
FT DOMAIN 19 56 NONHELICAL REGION (NC2).
FT DOMAIN 57 519 TRIPLE-HELICAL REGION.
FT DOMAIN 520 680 NONHELICAL REGION (NC1).
FT DOMAIN 545 680 C1Q.
FT CONFLICT 248 248 P -> L (IN REF. 3).
FT CONFLICT 286 286 A -> S (IN REF. 2).
FT CONFLICT 306 306 L -> F (IN REF. 3).
FT CONFLICT 417 417 T -> S (IN REF. 3).
FT CONFLICT 451 451 R -> K (IN REF. 4).
FT CONFLICT 500 500 H -> L (IN REF. 3).
FT CONFLICT 567 567 A -> C (IN REF. 3).
FT CONFLICT 569 569 I -> H (IN REF. 3).
FT CONFLICT 571 572 FD -> YI (IN REF. 3).
FT CONFLICT 635 635 Q -> T (IN REF. 3).
SQ SEQUENCE 680 AA; 56775 MW; FE984CA99AF708E2 CRC64;

Query Match      28.7%; Score 376; DB 1; Length 680;
Best Local Similarity 37.4%; Pred.No.1.3e-17;
Matches   91; Conservative 31; Mismatches 77; Indels 44; Gaps 9;

QY    36 GPRG-----PGPDG--APASVPPFPF-----PGAKGEVG---RRGRAGLRGGPGRPGRG 79
       |||||         |||         |||         |||         :|||         |
Db    440 GRPEGPIPTRGTGPVPVCGFGSGKDPGNPGAPGPAGIATKLNGLTGTPGPGPGPRG 499
       |||||         |||         |||         |||         :|||         |
QY    80 PRGERGRCPGPPGPGPCGVAP----AAGYPRI-----AFAVLRL 117
       |||||         |||         |||         |||         :|||         |
Db    500 HSSEGPLCLPGPPGPPGPGQAQMPVDGF IKAGQRPLSCGMPLVSANHGVTGMPVASFTVLS 559
       |||||         |||         |||         |||         :|||         |
QY    118 RPHEGYEV-LRGDDVVTVNVGNAYEAASGKFCTPMPCGYFFAYHVLVRGGDGTSWADLMK 176
       :||:::        ||:::        ||:::        ||:::        ||:::        |
Db    560 KAYPAVGAPIPDEILYNQQHYDPRSGLFITCKIPGIYFYFSYHVVK---GTRHWVGLYK 616
       :||:::        ||:::        ||:::        ||:::        ||:::        |
QY    177 NGQVRSATAQDAQNIDYASNIVLIHLDVDGVEVFIKLDGGKVGHGNTNKY--STPSGFI 234
       ||           ||           ||           ||           ||           |
Db    617 NCTPTMYTYDEVSKGYLDQAGSAIMELTENDQVWLQLPNAESNGLYSSEYVHSFSGFL 676
       ||           ||           ||           ||           ||           |
QY    235 IYP 237
       :||
Db    677 VAP 679
```

```
FT CONFLICT 464 464 A -> H (IN REF. 1).
FT CONFLICT 601 601 Y -> T (IN REF. 1).
FT CONFLICT 631 631 A -> G (IN REF. 1).
SQ SEQUENCE 744 AA; 73364 MW; 2BC1B0955DE2C9A3 CRC64;

Query Match
Best Local Similarity 37.2%; Pred. No. 1.4e-17; Length 744;
Matches 90; Conservative 31; Mismatches 77; Indels 44; Gaps 10;

QY 34 PHGPRG-PGPDGAPASVPPFPKAGKGVGRGKAGLRGPPGPPGPGPCGRGPPGP 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 508 PIGPPIPGKGEPLG-PGPGFPG-IGKPGVAGLHPPGKPGALGPGQPGLPGP 564
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 93 PG-PGPGGV-----APAAGYVPRIAF-YAGLRPH-----EGYEV----- 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 565 PGPPGPPAVMPTPPPGQGLVDMGLGIDGKPPHAYGAKKNGKNGSPAYEMPFAETA 624
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 126 -----LRFDVTVNNGNAYAAAGKTCMPGPGVFFAYHVLVRGGDGTSMWADLMKN 177
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 625 PFPVPGAPVKFNKLLNGRQNPOTGIFTCEVPGVYFFAYHVKCKG---NWVALFRN 681
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 178 GOVRSAAIQADQNDYASNVILHLDVGDVEFIKLDGKVKHGGNTNKV---STFSGFII 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 682 NEPVMTYDEYKKGFLDQAGSNVLLLRGDRVFLQMPSEQAAGLYAGQIVHSSFGYLL 741
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 YP 237
   ||
Db 742 YP 743

RESULT 8
CAJA_HUMAN STANDARD; PRT; 680 AA.
ID CA03692;
AC 003692;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(X) chain precursor.
GN COL10A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SQ SEQUENCE FROM N.A.
RX MEDLINE-92109659; PubMed-1764025;
RA Thomas J.T., Cresswell C.J., Rash B., Nicolai H., Jones T.,
RA Solomon E., Grant M.E., Boot-Handford R.P.;
RT "The human collagen X gene. Complete primary translated sequence and
RT chromosomal localization."
RL Biochem. J. 280:617-623(1991).
RN [2]
SQ SEQUENCE FROM N.A.
RX MEDLINE-93012005; PubMed-1397333;
RA Reichenberger E., Beier F., Luvall P., Olsen B.R., von der Mark K.,
RA Bertling W.M.;
RT "Genomic organization and full-length cDNA sequence of human collagen
RT X."
RL FEBS Lett. 311:305-310(1992).
RN [3]
SQ SEQUENCE FROM N.A.
RA Beier F., Lammi M.B., von der Mark K.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
SQ SEQUENCE FROM N.A.
RA Williams S.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
SQ SEQUENCE OF 52-680 FROM N.A.
RX MEDLINE-92267014; PubMed-1597271;
RA Apte S.S., Seldin M.F., Hayashi M., Olsen B.R.;
RT "Cloning of the human and mouse type X collagen genes and mapping of
RT the mouse type X collagen gene to chromosome 10.";
```

```
RL Eur. J. Biochem. 206:217-224(1992).
RN [6]
SQ SEQUENCE OF 561-666 FROM N.A.
RX MEDLINE-91243838; PubMed-2037056;
RA Apte S., Mattei M.G., Olsen B.R.;
RT "Cloning of human alpha 1(X) collagen DNA and localization of the
RT COL10A1 gene to the q21-q22 region of human chromosome 6."
RL FEBS Lett. 282:393-396(1991).
RN [7]
SQ SEQUENCE OF 547-655 FROM N.A.
RX MEDLINE-92077285; PubMed-1743401;
RA Reichenberger E., Aigner T., von der Mark K., Stoeb H., Bertling W.;
RT "In situ hybridization studies on the expression of type X collagen
RT in fetal human cartilage."
RL Dev. Biol. 148:562-572(1991).
RN [8]
REVIEW ON VARIANTS.
RX MEDLINE-9735959; PubMed-9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels."
RL Hum. Mutat. 9:300-315(1997).
RN [9]
VARIANTS SMCD ASP-598 AND PRO-614.
RX MEDLINE-94136476; PubMed-8304336;
RA Wallis G.A., Rash B., Sweetman W.A., Thomas J.T., Super M., Evans G.,
RA Grant M.E., Boot-Handford R.P.;
RT "Amino acid substitutions of conserved residues in the
RT carboxyl-terminal domain of the alpha 1(X) chain of type X collagen
RT occur in two unrelated families with metaphyseal chondrodysplasia
RT type Schmid."
RL Am. J. Hum. Genet. 54:169-178(1994).
RN [10]
VARIANT SMCD ARG-591.
RX MEDLINE-94272470; PubMed-8004099;
RA McIntosh I., Abbott M.H., Warman M.L., Olsen B.R., Francomano C.A.;
RT "Additional mutations of type X collagen confirm COL10A1 as the
RT Schmid metaphyseal chondrodysplasia locus."
RL Hum. Mol. Genet. 3:303-307(1994).
RN [11]
VARIANT SMCD VAL-618.
RX MEDLINE-95181449; PubMed-7876225;
RA Chan D., Cole W.G., Rogers J.G., Bateman J.F.;
RT "Type X collagen multimer assembly in vitro is prevented by a Gly618
RT to Val mutation in the alpha 1(X) NC1 domain resulting in Schmid
RT metaphyseal chondrodysplasia."
RL J. Biol. Chem. 270:4558-4562(1995).
RN [12]
VARIANTS SMCD ARG-545; GLU-595; HIS-597; LYS-617; ARG-644 AND GLY-648.
RX MEDLINE-95331767; PubMed-7607655;
RA Bonaventure J., Chaminade F., Maroteaux P.;
RT "Mutations in three subdomains of the carboxy-terminal region of
RT collagen type X account for most of the Schmid metaphyseal
RT dysplasias."
RL Hum. Genet. 96:58-64(1995).
RN [13]
VARIANT SMCD PRO-600.
RX MEDLINE-96375754; PubMed-8782043;
RA Wallis G.A., Rash B., Sykes B., Bonaventure J., Maroteaux P.,
RA Zabel B., Wynne-Davies R., Grant M.E., Boot-Handford R.P.;
RT "Mutations within the gene encoding the alpha 1(X) chain of type X
RT collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but
RT not several other forms of metaphyseal chondrodysplasia."
RL J. Med. Genet. 33:450-457(1996).
RN [14]
VARIANTS SMCD GLU-18 AND ARG-18.
RX MEDLINE-97220591; PubMed-9067753;
RA Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.;
RT "Mutations in the N-terminal globular domain of the type X collagen
RT gene (COL10A1) in patients with Schmid metaphyseal
RT chondrodysplasia."
RL Hum. Mutat. 9:131-135(1997).
```


615 VMTYDEYIKGYLDQASGAVIDLTDNQVLMQLFNAGSNGLYSPEYVHSSFGSLVAP 673

RESULT 11

CLQB_HUMAN
ID CLQB_HUMAN STANDARD; PRT; 251 AA.
AC P02746; Q96H17;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement C1q subcomponent, B chain precursor.
GN CLQB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=86076906; PubMed=3000358;
RA Reid K.B.M.;
RT "Molecular cloning and characterization of the complementary DNA and
RT gene coding for the B-chain of subcomponent C1q of the human
RT complement system.";
RL Biochem. J. 231:729-735(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 26-133.
RX MEDLINE=80020137; PubMed=486087;
RA Reid K.B.M.;
RT "Complete amino acid sequences of the three collagen-like regions
RT present in subcomponent C1q of the first component of human
RT complement.";
RL Biochem. J. 179:367-371(1979).
RN [4]
RP SEQUENCE OF 26-193.
RX MEDLINE=79041552; PubMed=708376;
RA Reid K.B.M., Thompson E.O.P.;
RT "Amino acid sequence of the N-terminal 108 amino acid residues of the
RT B chain of subcomponent C1q of the first component of human
RT complement.";
RL Biochem. J. 173:863-868(1978).
RN [5]
RP SEQUENCE OF 134-251.
RX MEDLINE=82283890; PubMed=6981411;
RA Reid K.B.M., Gagnon J., Frampton J.;
RT "Completion of the amino acid sequences of the A and B chains of
RT subcomponent C1q of the first component of human complement.";
RL Biochem. J. 203:559-569(1982).
RN [6]
RP SEQUENCE OF 224-251 FROM N.A.
RC Tissue=Liver;
RX MEDLINE=85038855; PubMed=6208566;
RA Reid K.B.M., Bentley D.R., Wood K.J.;
RT "Cloning and characterization of the complementary DNA for the B
RT chain of normal human serum C1q.";
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 306:345-354(1984).
RN [7]
RP REVIEW OF C1Q DEFICIENCY.
RX MEDLINE=98450587; PubMed=9777412;
RA Petry F.;
RT "Molecular basis of hereditary C1q deficiency.";
RL Immunobiology 199:286-294(1998).
CC -!- FUNCTION: C1Q ASSOCIATES WITH THE PROENZYMES C1R AND C1S TO YIELD
CC C1, THE FIRST COMPONENT OF THE SERUM COMPLEMENT SYSTEM. THE
CC COLLAGEN-LIKE REGIONS OF C1Q INTERACT WITH THE CA(2+)-DEPENDENT
CC C1R(2)C1S(2) PROENZYME COMPLEX, AND EFFICIENT ACTIVATION OF C1
CC TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE
CC FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.

CC -!- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R
CC AND S IN THE MOLAR RATIO OF 1:2:2. C1Q SUBCOMPONENT IS COMPOSED
CC OF NINE SUBUNITS, SIX OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE
CC A AND B CHAINS, AND THREE OF WHICH ARE DISULFIDE-LINKED DIMERS OF
CC THE C CHAIN.
CC -!- PTM: O-LINKED GLYCANS CONSIST OF GLC-GAL DISACCHARIDES.
CC -!- DISEASE: DEFECTS IN C1QB ARE A CAUSE OF C1Q DEFICIENCY. IT IS A
CC RARE GENETIC DISORDER WHICH IS ASSOCIATED WITH RECURRENT
CC INFECTIONS AND A HIGH PREVALENCE OF LUPUS ERYTHEMATOSUS-LIKE
CC SYMPTOMS. IT IS CHARACTERIZED BY A LOSS OF ACTIVATION OF THE
CC COMPLEMENT CLASSICAL PATHWAY.
CC -!- SIMILARITY: CONTAINS 1 COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X03084; CAA26880.1; -
DR EMBL; BC008983; RAH08983.1; ALT_INIT.
DR EMBL; M36278; AAC41692.1; -
DR PIR; A03206; C1HUQB.
DR PIR; B23422; B23422.
DR Genew; HGNC:1242; C1QB.
DR MIM: 120570; -
DR InterPro; IPR001073; C1q.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; C1q; 1.
DR Pfam; PF01391; Collagen; 1.
DR PRINTS; PR00007; COMPLENNTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1q; 1.
KW Complement pathway; Plasma; Hydroxylation; Glycoprotein; Collagen;
KW Repeat; Signal; Disease mutation.
FT SIGNAL 1 25
FT CHAIN 26 251
FT DOMAIN 29 112
FT DOMAIN 113 251
FT MOD_RES 26 26 PYRROLIDONE CARBOXYLIC ACID.
FT DISULFID 29 29 INTERCHAIN (WITH C-26 IN CHAIN A).
FT MOD_RES 33 33 HYDROXYLATION.
FT MOD_RES 36 36 HYDROXYLATION.
FT MOD_RES 39 39 HYDROXYLATION.
FT MOD_RES 42 42 HYDROXYLATION.
FT MOD_RES 51 51 HYDROXYLATION.
FT MOD_RES 54 54 HYDROXYLATION.
FT MOD_RES 57 57 HYDROXYLATION.
FT CARBOHYD 57 57 O-LINKED (GAL. .).
FT MOD_RES 60 60 HYDROXYLATION.
FT CARBOHYD 60 60 O-LINKED (GAL. .).
FT MOD_RES 63 63 HYDROXYLATION.
FT MOD_RES 75 75 HYDROXYLATION.
FT MOD_RES 81 81 HYDROXYLATION.
FT MOD_RES 84 84 HYDROXYLATION.
FT MOD_RES 90 90 HYDROXYLATION.
FT MOD_RES 96 96 HYDROXYLATION.
FT CARBOHYD 96 96 O-LINKED (GAL. .).
FT MOD_RES 99 99 HYDROXYLATION.
FT MOD_RES 102 102 HYDROXYLATION.
FT MOD_RES 105 105 HYDROXYLATION.
FT MOD_RES 108 108 HYDROXYLATION.
FT CARBOHYD 108 108 O-LINKED (GAL. .).
FT VARIANT 40 40 G -> D (IN C1Q DEFICIENCY).
FT FTIG-Var_008541.
FT CONFLICT 26 26 Q -> E (IN REF. 3).
FT CONFLICT 83 83 N -> D (IN REF. 3).
FT CONFLICT 98 98 G -> P (IN REF. 3 AND 4).
FT SEQUENCE 251 AA; 26459 MW; 78C5752E267A0EF7 CRC64;

[illegible]

[4] SEQUENCE OF 29-122.
RX MEDLINE=80020137; PubMed=486087;
RA Reid K.B.M.;
RT "Complete amino acid sequences of the three collagen-like regions
RL present in subcomponent Clq of the first component of human
RN Biochem. J. 179:367-371(1979).
[5]
RP REVIEW OF C1Q DEFICIENCY.
RX MEDLINE=98450587; PubMed=9777412;
RA Petry F.;
RT "Molecular basis of hereditary C1q deficiency.";
RL Immunobiology 199:286-294(1998).
CC -!- FUNCTION: C1Q ASSOCIATES WITH THE PROENZYMES C1R AND C1S TO YIELD
CC COLLAGEN-LIKE REGIONS OF C1Q INTERACT WITH THE CA(2+)-DEPENDENT
CC C1R(2)C1S(2) PROENZYME COMPLEX, AND EFFICIENT ACTIVATION OF C1
CC TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE
CC FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.
CC -!- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R
CC AND S IN THE MOLAR RATIO OF 1:2:2. C1Q SUBCOMPONENT IS COMPOSED
CC OF NINE SUBUNITS; SIX OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE
CC A AND B CHAINS, AND THREE OF WHICH ARE DISULFIDE-LINKED DIMERS OF
CC THE C CHAIN.
CC -!- PTM: O-LINKED GLYCANS CONSIST OF GLC-GAL DISACCHARIDES.
CC -!- DISEASE: DEFECTS IN C1QG ARE A CAUSE OF C1Q DEFICIENCY. IT IS A
CC RARE GENETIC DISORDER WHICH IS ASSOCIATED WITH RECURRENT
CC INFECTIONS AND A HIGH PREVALENCE OF LUPUS ERYTHEMATOSUS-LIKE
CC SYMPTOMS. IT IS CHARACTERIZED BY A LOSS OF ACTIVATION OF THE
CC COMPLEMENT CLASSICAL PATHWAY.
CC -!- SIMILARITY: CONTAINS 1 COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).

EMBL; AK057792; BAB71575.1; -
DR EMBL; BC009016; AAH09016.1; -
DR PIR; A03207; C1HUQC
DR PIR; S14351; S14351.
DR Genew; HGNC:1245; C1QG.
DR MIM; 120575; -
DR InterPro; IPR001073; C1q.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; C1q; 1.
DR Pfam; PF01391; Collagen; 1.
DR PRINTS; PR00007; COMPLENNTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
DR Complement pathway; Plasma; Hydroxylation; Glycoprotein; Collagen;
KW Repeat; Signal; Disease mutation.
FT SIGNAL 1 28
FT CHAIN 29 245 COMPLEMENT C1Q SUBCOMPONENT, C CHAIN.
FT DOMAIN 31 112 COLLAGEN-LIKE.
FT DOMAIN 113 245 C1Q.
FT DISULFID 32 32 INTERCHAIN (WITH OTHER C CHAIN).
FT MOD_RES 36 36 HYDROXYLATION.
FT MOD_RES 39 39 HYDROXYLATION.
FT MOD_RES 42 42 HYDROXYLATION.
FT MOD_RES 45 45 HYDROXYLATION.
FT MOD_RES 54 54 HYDROXYLATION.
FT MOD_RES 57 57 HYDROXYLATION.
FT MOD_RES 63 63 HYDROXYLATION.
FT MOD_RES 66 66 HYDROXYLATION.
FT MOD_RES 71 71 HYDROXYLATION.
FT MOD_RES 75 75 HYDROXYLATION.
FT CARBOHYD 75 75 O-LINKED (GAL.).

PFAM: PF01391; Collagen; 2.
DR PRINTS; PR00007; COMPLENNTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
DR KW Collagen; Signal.
FT FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 285 COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-
RELATED PROTEIN 2.
FT DOMAIN 40 141 COLLAGEN-LIKE.
FT FT DOMAIN 143 285 C1Q.
SQ SEQUENCE 285 AA; 29952 MW; 7E31FF9F868DAEDFA CRC64;

Query Watch 25.0%; Score 328; DB 1; Length 285;
Best Local Similarity 33.6%; Pred. No. 7.2e-15;
Matches 86; Conservative 37; Mismatches 83; Indels 50; Gaps 10;

QY 26 GRCMVCDPHGRG-PGPDGAPA-----SYPPPPGA 56
Db 30 GSFQLVCSLPGGPGPPGAPGSGMMRGFMFGKCGQDGDHSDGSSEEGPGRGTGN 89
QY 57 KGEVGRKGAGLRGPGRGPPGPRGPPGPPGPG-----PGGVAAAGAYVP 108
Db 90 RGKPGPKKAGAIGRAGRPGKVGTPGKHGTGFKGKPGKGLPCSCGSGHT- 148
QY 109 RTAFVAGLR--PHEGYEVLRDDVVTVNGVNAYEAASGKTFCMPMGVYFYAHVLMRGD 166
Db 149 KSASFVAVTKSPYRRPLI-KFDKILMEGHYNASSGKFVCVPGVIYYTYDTTLA--- 204
QY 167 GTSMWADLMKNGOVRSALADADQ-NVDYASNVLHLVDGEVFVKLOGGKVHGNTN 225
Db 205 NKHLAIGLVHNGQYRIETF--DANTGNHDVASGSTLTALKQGDDEVWLQIFYSEQNLGYD 262
QY 226 KY---STFSGFIIYPD 238
Db 263 PYWTDLSLTGFLIYAD 278

RESULT 14
C1QC_HUMAN STANDARD; PRT; 245 AA.
ID C1QC_HUMAN AC P02747; Q96DL2; Q96H05;
AC 21-JUL-1986 (Rel. 01, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement C1q subcomponent, C chain precursor.
GN C1Q OR C1QC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
OX [1]
SEQUENCE FROM N.A.
RC TISSUE=Monocytes;
RP MEDLINE=91174759; PubMed=1706597;
RX Sellar G.C., Blake D.J., Reid K.B.M.;
RT "Characterization and organization of the genes encoding the A-, B-
and C-chains of human complement subcomponent C1q. The complete
derived amino acid sequence of human C1q";
RL Biochem. J. 274:481-490(1991).
RN [2]
SEQUENCE FROM N.A.
RP RP TISSUE=Cerebellum;
RC Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Suzuki Y.,
RA Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,
RA Sugiyama T., Irie R., Osuki T., Sato H., Nishikawa T., Sugiyama A.,
RA Kawakami B., Nagai K., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RP RP TISSUE=Brain;
RC Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

Decreased adiponectin plasma levels are associated with obesity
insulin resistance, and diabetes type 2.
-!- PHARMACEUTICAL: Adiponectin might be used in the treatment of
diabetes type 2 and insulin resistance.
-!- SIMILARITY: CONTAINS 1 COLLAGENOUS DOMAIN.
-!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D45371; BAA08227.1; ..
DR EMBL; AB012165; BAA86716.1; ..
DR EMBL; AB012164; BAA86716.1; JOINED.
DR EMBL; AJ131460; CAB52413.1; ..
DR EMBL; AJ131461; CAB52413.1; JOINED.
DR MIM; 605441; ..
DR InterPro; IPR001073; Clq.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; Clq; 1.
DR Pfam; PF01391; Collagen; 1.
DR PRINTS; PR000007; COMPLEMENTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
KW Hormone; Collagen; Signal; Repeat; Hydroxylation; Plasma;
KW Polymorphism; Disease mutation; Obesity; Diabetes mellitus.
FT SIGNAL 1 14
FT CHAIN 15 244
FT DOMAIN 42 107
FT DOMAIN 108 244
FT DISULFID 36 36
FT MOD_RES 44 44
FT MOD_RES 47 47
FT MOD_RES 53 53
FT MOD_RES 62 62
FT MOD_RES 71 71
FT MOD_RES 76 76
FT MOD_RES 86 86
FT MOD_RES 95 95
FT MOD_RES 104 104
FT VARIANT 84 84
FT VARIANT 112 112
FT VARIANT 117 117
FT VARIANT 164 164
FT VARIANT 221 221
FT VARIANT 241 241
SQ SEQUENCE 244 AA; 26414 MW; 64D8C6C1204B1018 CRC64;

Query Match 24.3%; Score 318; DB 1; Length 244;
Best Local Similarity 34.3%; Pred. No. 2.7e-14;
Matches 92; Conservative 36; Mismatches 82; Indels 58; Gaps 12;

QY 3 LLLVAIPLLVHSSRGPAHYEMGLGRCRMVCDPHGPRGP-----PDGA----PASVPPFP 53
Db 1 MLLIGAVLLLLAL---PCHDQ-----ETTQPGVLLPLPKGACTGWMAGIPGHP 47

QY 54 -----PGAKGVGRGKAGLRGPPGPPGPPGPPGPPGPPG-----PGPG 98
Db 48 GHNGAPGRDGRDTPGEGKGGDPLIGPKDIGETGVPGAEGPRGPGIQRKGEPEG 107

QY 99 GVAPAGVVPRIATYAGLRPHGEYEV-----LRFDDVVTVNNGNAYEAASKFTCPMPGV 153
Db 108 -----AYVRSASFVGL-----ETYVTIPNMPIRFTKIFYNQNHGDSGKFCNIPGL 157

OY 154 YFFAYHVLMRGGDGTSMWADLMKNGQVRASAIQAQDQNDQYASNSVILHLVDVGDEVF 213
Db 158 YFFAYHITVYMKD---VKVSLFKKDKKAMLFYDQYQENNVDQASGVLLHLEVGQVWLQ 214
OY 214 LDG-GKVHG--GNTNKYSTFGFIYPD 238
Db 215 VYGEGRNGLYADNDNDSTFTGFLLYHD 242

Search completed: June 2, 2003, 16:45:59
Job time : 23 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query	§			ID	Description
			Match	Length	DB		
1	327	24.9	294	4	US-09-188-930-294	Sequence 294, Appl	
2	318	24.3	244	2	US-08-463-911-7	Sequence 7, Appli	
3	318	24.3	244	4	US-09-140-804-3	Sequence 3, Appli	
4	318	24.3	244	4	US-09-336-536-20	Sequence 20, Appl	
5	318	24.3	244	4	US-08-530-423-1	Sequence 1, Appli	
6	317.5	24.2	247	2	US-08-463-911-2	Sequence 2, Appli	
7	316.5	24.1	231	4	US-09-530-423-2	Sequence 2, Appli	
8	316	24.1	246	2	US-08-463-911-4	Sequence 4, Appli	
9	314	24.0	245	4	US-09-140-804-4	Sequence 4, Appli	
10	309.5	23.6	247	4	US-09-140-804-8	Sequence 8, Appli	
11	309.5	23.6	247	4	US-09-118-408-3	Sequence 3, Appli	
12	309.5	23.6	247	4	US-09-508-855-3	Sequence 3, Appli	
13	300	22.9	228	4	US-09-336-536-4	Sequence 4, Appli	
14	300	22.9	243	4	US-09-188-930-295	Sequence 2, Appli	
15	300	22.9	243	4	US-09-140-804-2	Sequence 2, Appli	
16	300	22.9	243	4	US-09-336-536-3	Sequence 3, Appli	
17	299	22.8	228	4	US-09-336-536-11	Sequence 11, Appl	
18	299	22.8	243	4	US-09-336-536-10	Sequence 10, Appl	
19	268.5	20.5	215	4	US-09-140-804-5	Sequence 5, Appli	
20	257	19.6	236	4	US-09-140-804-6	Sequence 6, Appli	
21	254.5	19.4	185	2	US-08-463-911-3	Sequence 3, Appli	
22	250	19.1	423	1	US-08-383-744-2	Sequence 2, Appli	
23	250	19.1	423	2	US-08-999-336-2	Sequence 2, Appli	
24	250	19.1	423	5	PCT-US96-01427-2	Sequence 2, Appli	
25	235.5	18.0	330	1	US-08-642-255-32	Sequence 32, Appl	
26	235.5	18.0	408	1	US-07-609-716-65	Sequence 65, Appl	
27	235.5	18.0	408	4	US-08-475-411A-65	Sequence 65, Appl	

; Sequence 7, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/463,911
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI95-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-463-911-7

Query Match 24.3%; Score 318; DB 2; Length 244;
Best Local Similarity 34.3%; Pred. No. 6e-20;
Matches 92; Conservative 36; Mismatches 82; Indels 58; Gaps 12;
QY 3 LLLVAIPLLVHSSRGPAHYEMLGRCRMVCDPHGPRGP-----PDGA-----PASVPPPP 53
DB 1 MLLLGAVLLALL--PGHDQ-----ETTQGGVLLPLPKGACTGWMAGIPGHP 47
QY 54 -----PGAKEVGRCKAGLRGPPGPPGPRGPPGPPGPPG-----PGPG 98
DB 48 GHNGAPGRDGRDGTGCEKGEKGDPLGPKDGTGTVGPAEGPRGFPGIGRKGPEGEG 107
QY 99 GVAPAGYVPRIAFYAGLRPHEGYEV-----LRFDVVTVNMGNAVYEAASGKFTCPMPGV 153
DB 154 YFAYHVMRGDGTSMWADLMKNGQVRSATQADQNDYASNSVILHLVDGVEFTK 213
DB 158 YFAYHITVYMKD---VKVSLFKKDKAMLTFTDQYQENNVDQASGVLLHLEVDQVWLQ 214
QY 214 LDG-GKVGH--GNTNKYSTFSGFIIPD 238
DB 215 VYGEGERNGLYADNDNDSTFTGFLLYHD 242

RESULT 3
US-09-140-804-3
; Sequence 3, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49

; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-140-804-3
Query Match 24.3%; Score 318; DB 4; Length 244;
Best Local Similarity 34.3%; Pred. No. 6e-20;
Matches 92; Conservative 36; Mismatches 82; Indels 58; Gaps 12;
QY 3 LLLVAIPLLVHSSRGPAHYEMLGRCRMVCDPHGPRGP-----PDGA-----PASVPPPP 53
DB 1 MLLLGAVLLALL--PGHDQ-----ETTQGGVLLPLPKGACTGWMAGIPGHP 47
QY 54 -----PGAKEVGRCKAGLRGPPGPPGPRGPPGPPGPPG-----PGPG 98
DB 48 GHNGAPGRDGRDGTGCEKGEKGDPLGPKDGTGTVGPAEGPRGFPGIGRKGPEGEG 107
QY 99 GVAPAGYVPRIAFYAGLRPHEGYEV-----LRFDVVTVNMGNAVYEAASGKFTCPMPGV 153
DB 154 YFAYHVMRGDGTSMWADLMKNGQVRSATQADQNDYASNSVILHLVDGVEFTK 213
DB 158 YFAYHITVYMKD---VKVSLFKKDKAMLTFTDQYQENNVDQASGVLLHLEVDQVWLQ 214
QY 214 LDG-GKVGH--GNTNKYSTFSGFIIPD 238
DB 215 VYGEGERNGLYADNDNDSTFTGFLLYHD 242

RESULT 4
US-09-336-536-20
; Sequence 20, Application US/09336536
; Patent No. 6405884
; GENERAL INFORMATION:
; APPLICANT: Leiby, K.
; APPLICANT: McKay, C.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-144
; CURRENT APPLICATION NUMBER: US/09/336,536
; CURRENT FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-336-536-20

Query Match 24.3%; Score 318; DB 4; Length 244;
Best Local Similarity 34.3%; Pred. No. 6e-20;
Matches 92; Conservative 36; Mismatches 82; Indels 58; Gaps 12;
QY 3 LLLVAIPLLVHSSRGPAHYEMLGRCRMVCDPHGPRGP-----PDGA-----PASVPPPP 53
DB 1 MLLLGAVLLALL--PGHDQ-----ETTQGGVLLPLPKGACTGWMAGIPGHP 47
QY 54 -----PGAKEVGRCKAGLRGPPGPPGPRGPPGPPGPPG-----PGPG 98
DB 48 GHNGAPGRDGRDGTGCEKGEKGDPLGPKDGTGTVGPAEGPRGFPGIGRKGPEGEG 107
QY 99 GVAPAGYVPRIAFYAGLRPHEGYEV-----LRFDVVTVNMGNAVYEAASGKFTCPMPGV 153
DB 154 YFAYHVMRGDGTSMWADLMKNGQVRSATQADQNDYASNSVILHLVDGVEFTK 213
DB 158 YFAYHITVYMKD---VKVSLFKKDKAMLTFTDQYQENNVDQASGVLLHLEVDQVWLQ 214
QY 214 LDG-GKVGH--GNTNKYSTFSGFIIPD 238
DB 215 VYGEGERNGLYADNDNDSTFTGFLLYHD 242

Matches 84; Conservative 32; Mismatches 72; Indels 45; Gaps 10;

```
QY 38 RGP-----PQCA-----PASVPPF-----PGAKGVRRGKAGLRGPPGPPGPPGPE 83
Db 10 QGPGVLLPLPKACTGWAIGIPCHGNCAPGRDGTTPGKEGKGDGLGPKGDIGE 69
QY 84 PGRPGPPGPG-----PGPGVAPAAGYVPRIAFYAGLRPHGEYEV-----LRF 128
Db 70 TGVPAEGPRGPGIQRKGEPEG-----AYVRSAPSVGL-----ETVVTIPNPIRF 119
QY 129 DDVTVNNGNAYEASGKTCMPGYYFFAYHVLMRGGDGTSMWADLMKNGQVRASIAQD 188
Db 120 TKIFYNQNHQDGTGKHCNIPGYFYFAYHITVYMKD---VKVSLFKKDKAMLFYDQY 176
QY 189 ADQNTDYASNSVILHLDVGEYFIKLDG-GKVHG--GNTNKYSTFSGFIYPD 238
Db 177 QENNVDAQSGSVLLHLEVDQWLVQYGEGERGLYADNDNDSTFTGFLLYHD 229
```

RESULT 8

```
US-08-463-911-4
; Sequence 4, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI95-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-463-911-4
```

Query Match 24.1%; Score 316; DB 2; Length 246;
Best Local Similarity 31.9%; Pred. No. 9e-20;
Matches 84; Conservative 38; Mismatches 85; Indels 56; Gaps 9;

```
QY 1 WYLLLVLAIPLLVHSSRGPAHYEMLGRWCRCVDPGPRGPGD-----APASVPPFP-- 53
Db 15 LLLLFLLALPLRSASAG--CYGIPGM-----PGMPGAPGKDGHDLGQGPKEGPIAV 66
QY 54 PGAKGEVRRGKAGLRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 112
Db 67 PCTQGPKGQKGPFGPHGRKGNPRGTSGLPDGPPGPPGPPGPPGPPGPPGPPGPPGPP 113
```

```
QY 113 YAGLRPH-----EGYEVLRFDVVTVNMGVNAEASGKFTCPMPGVYFFAY 158
Db 114 --RYQKHQSFTVTRQTOYPEANALVRFNSVTVNPOGHYNPSTGKETCEVGLYFYV 171
QY 159 HVLMRGGDGTSMWADLMKNGQVRASIAQADADQNYD---YASNSVILHLDVGEYFIKLD 215
Db 172 Y-----TSTANLCVHLNLTARVASFCDHMFNSKQVSSGGALLRLQRGDEVWLSVN 223
QY 216 GGVHGGNTNKYSTFSGFIYPD 238
Db 224 DINGWVGIBGSNSVFSGFLFPD 246
```

RESULT 9

```
US-09-140-804-4
; Sequence 4, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-140-804-4
```

Query Match 24.0%; Score 314; DB 4; Length 245;

Best Local Similarity 36.6%; Pred. No. 1.3e-19;
Matches 87; Conservative 23; Mismatches 80; Indels 48; Gaps 11;

```
QY 28 CRMVCDPHGPRG-----PGPDGAPASVPPPPPGAKGEVG-----RRGKAGLRGPPGPPG 76
Db 26 CRA---PDGKKGEAGRPGRRR-----PGLKGEQGEPPGAPGIRTIQGLKGDQGE 74
QY 77 PRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 129
Db 75 PSNPGKVGYPGSPGLGARGIPGKGTGSPGNKQPRPAFSAIRRNPPMGGNVVIFD 134
QY 130 DVTVNMGVNAEASGKFTCPMPGVYFFAYHVLMRGGDGTSMWADLM-----KNGOVRASA 184
Db 135 TVITNQEEPYQNHSGRFVCTVPGYIFTFOVL-----SOMEICLSIVSSSRGQVRRSL 187
QY 185 IAQDADQN---YDYASNSVILHLDVGEYFIKLDGKGVH---GGNTNKYSTFSGFIY 237
Db 188 GFCDITNKGLFQVSGGWLQLQQGDQVWVKDPKKGHIYQSEAD--SVFSGFLFP 243
```

RESULT 10

```
US-09-140-804-8
; Sequence 8, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 247
; TYPE: PRT
```


Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	666	50.8	268	9	US-09-998-582-2		Sequence 2, Appli
2	662.5	50.5	250	9	US-09-998-563-2		Sequence 2, Appli
3	376	28.7	744	9	US-10-171-311-40		Sequence 40, Appli
4	345.5	26.4	638	9	US-10-001-887-108		Sequence 108, App
5	345.5	26.4	703	9	US-10-219-449-4		Sequence 4, Appli
6	345.5	26.4	717	9	US-10-219-449-2		Sequence 2, Appli
7	327	24.9	285	9	US-09-866-050A-382		Sequence 382, App
8	327	24.9	294	9	US-09-866-050A-294		Sequence 294, App
9	319.5	24.4	289	9	US-09-866-050A-630		Sequence 630, App
10	318	24.3	244	9	US-09-911-1768-48		Sequence 48, Appl
11	318	24.3	244	9	US-10-180-762-51		Sequence 51, Appl
12	318	24.3	244	10	US-09-776-976-6		Sequence 6, Appl
13	318	24.3	244	10	US-09-758-055-6		Sequence 6, Appl
14	318	24.3	244	10	US-09-909-547-6		Sequence 6, Appl
15	317.5	24.2	247	10	US-09-776-976-4		Sequence 4, Appl
16	317.5	24.2	247	10	US-09-758-055-4		Sequence 4, Appl
17	317.5	24.2	247	10	US-09-909-547-4		Sequence 4, Appl
18	316	24.1	245	9	US-09-800-729-116		Sequence 116, Ap
19	314	24.0	245	9	US-09-911-1768-49		Sequence 49, Appl

[illegible]

```

RESULT 5
US-10-219-449-4
; Sequence 4, Application US/10219449
; Publication No. US20030077820A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan (Sean)
; APPLICANT: Xie, Qiongsu
; APPLICANT: Hu, Yi
; TITLE OF INVENTION: No. US20030077820A1 Human Collagen Proteins and Polynucleotides
; FILE REFERENCE: LEX-0363-USA
; CURRENT APPLICATION NUMBER: US/10/219,449
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: US 60/312,300
; PRIOR FILING DATE: 2001-08-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 703
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-219-449-4

```

Query Match	26.4%	Score	345.5;	DB	9;	Length	703;
Best Local Similarity	35.6%;	Pred.	No. 3e-16;				
Matches	94;	Conservative	30;	Mismatches	77;	Indels	63;
Gaps	12;						
QY	17	RGAHYEMLCRMCWCDPHGRPGP-GPDGAPA-SVPFPFPCAKEVGERRKAGLGRGPPGP	74				
Db	459	RGFSGIPGL-----QCAGPIGPQGLGKLGEPLGPQCE-CRAGEPTAGTGP	508				
QY	75	PGRGRPPEGRCPGPCPGPGGGVAPAGAAYVPRIAFYAGLRPRHGEYE-----	124				
Db	509	PGVPGSPGITGPPGPGPPGP-PG--APGAFDETGI---AGLHLNPGSGEVAVLGKGKGP	562				
QY	125	-----VLRFDDVTNVNGNAVEAASGKTFCMPGVYFF	156				
Db	563	QFGLGELSAHATPAFTAVLTSPFASCMPVKFEDRTLYNGHSYNPATGICTCPGVGYVF	622				
QY	157	AYHVLMRGGDGTSMADLMKNQVRASAIQAODQNT-DYASNVIHLVDGDEVFKLD	215				
Db	623	AIVHHVK---GTNNVALYKN-NVPATYYDEKKGYLDQASGAVQLRLPNQDVVQMP	678				
QY	216	GKVHGQNTNKY--STFGSFIYP	237				
Db	679	SDOANGLYSTEYIHTSSFSGLLCP	702				

```

RESULT 6
US-10-219-449-2
: Sequence 2, Application US/10219449
: Publication No. US20030077820A1
: GENERAL INFORMATION:
: APPLICANT: Yu, Xuanchuan (Sean)
: APPLICANT: Xie, Qiongsu
: APPLICANT: Hu, Yi
: TITLE OF INVENTION: No. US20030077820A1el Human Collagen Proteins and Polynucleotides
: FILE REFERENCE: LEX-0363-USA
: CURRENT APPLICATION NUMBER: US/10/219,449
: CURRENT FILING DATE: 2002-08-14
: PRIOR APPLICATION NUMBER: US 60/312,300
: PRIOR FILING DATE: 2001-08-14
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FastSEQ for Windows Version 4.0

```

```

; SEQ ID NO 2
; LENGTH: 717
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-219-449-2

```

Query Match	26.4%;	Score 345.5;	DB 9;	Length 717;
Best Local Similarity	35.6%;	Pred. NO. 3.1e-16;		
Matches 94;	Conservative 30;	Mismatches 77;	Indels 63;	Gaps 12;

Qy	17	RGPAHYEMLGRCRVCDPHGDRGP--GPDGAPA-SVPPPPGAKGVGRRGKAGLRGPPGP	74
Db	473	RGSPGIPGL-----QGAPGIPGGLPELKGEPGLPDPGPG--GRAGEPTAGTGP	522
Qy	75	PGPRGPGEPRGPDPGPGPGGVAAGVVPRIAEVAGLRRRHEGYE-----	124
Db	523	PGVPGSPGITGPPGPPGPPGPG--PG--APGAFDETCI---AGLHLPNGGVEGAVLGKGGK	576
Qy	125	-----VLRFDVVTVNGVNAEAAASKFTCPMPGVYVF	156
Db	577	QFGELGELSAHATPAFTAVLTSPFPASGMFVFDRTLYNGHSGYNPATGFTCPVGGVYF	636
Qy	157	AYHVLMGGDGTSMWDLMMKQVRSAAIADQNY--DYASNSVILHLVDGVDFYIKLD	215
Db	637	AYHVHK---GTNVWVALYKN-NVPATVYDEYKKGYLDQASGGAVLQLRPNDQVWVQMP	692
Qy	216	GGKVHGGNTNKY--STFSGFIIYP	237
Db	693	SDQANGLYSTEYIHSSFSGFLICP	716

```

RESULT 7
US-09-866-050A-382
; Sequence 382, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 382
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Rat
US-09-866-050A-382

```

	Query Match	24.98;	Score 327;	DB 9;	Length 285;
	Best Local Similarity	32.7%;	Pred. No. 2.3e-15;		
	Matches 86;	Conservative 33;	Mismatches 80;	Indels 64;	Gaps 9
Qy	26	GRCRMVDPHGPRG-PGPDGAPASV-----	-----pppfPGA	56	
		: : : :			
Db	30	GGPOLVCSLPGPQGPPGPAGPSSGMVRMGFFPKDGDGDGRDSDGEEGPPCRTGN	89		
		: : : :			
Qy	57	KGEVRRGKAGLRCPGPCPPGPGRPGPCPPG---CGGCVAAAGVVPRIA	111		
		: : : :			
Db	90	RKGQPGKKAGATRAGRPCKGVSPGKKHIGTGGKPKKKKEPGLPGPCS	142		
		: : : :			
Qy	112	FYAGLRPHRGYEY-----LRFDVVTVNGVAYEAAASKFTCPMPGVVFYFAYH	159		
		: : : :			
Db	143	--CGSSRAKSAFSVSTKSPRRLPIKFCDILANEGGHYNASSGKFCVSPGIYYFYTD	200		
		: : : :			
Qy	160	VLMRGDGTSWADLMKNQVGRASIAQDADQ--NYDYASNVSIIHLDDYGFVKLDGGK	218		
		: : : :			

Result No.	Score	Query %			DB	ID	Description
		Match	Length	Time			
1	389	29.7	743	1	S23779	collagen alpha 1(V	
2	388.5	29.6	744	1	A34246	collagen alpha 1(V	
3	383.5	29.3	744	1	S23298	collagen alpha 1(V	
4	376	28.7	680	2	S12126	collagen alpha 1(X	
5	376	28.7	744	2	S15435	collagen alpha 1(V	
6	369.5	28.2	674	2	S23297	collagen alpha 1(X	
7	362.5	27.7	680	1	CGHUIID	collagen alpha 1(X	
8	360	27.5	674	2	S13301	collagen alpha 1(X	
9	352	26.8	253	1	C1HUQB	complement subcomp	
10	345	26.3	635	2	A57131	collagen alpha 2(V	
11	322	24.6	245	1	C1HUQC	complement subcomp	
12	318	24.3	244	2	JC4708	gelatin-binding 28	
13	316	24.1	246	2	S29328	complement subcomp	
14	316	24.1	253	2	S49158	complement subcomp	
15	314	24.0	245	1	C1HUQA	complement subcomp	
16	305.5	23.3	253	2	I49560	complement Clq B c	
17	302	23.0	245	2	S19018	complement subcomp	
18	300	22.9	219	2	T14782	hypothetical prote	
19	268.5	20.5	215	2	B48150	hibernation-relate	
20	261.5	19.9	215	2	C48150	hibernation-relate	
21	260.5	19.9	992	2	T08772	hypothetical prote	
22	250	19.1	423	2	A55797	collagen precursor	
23	243	18.5	224	2	A60032	cerebellin-like gl	
24	229	17.5	193	2	A37873	cerebellin precurs	
25	222.5	17.0	196	2	A48150	hibernation-relate	
26	221.5	16.9	464	2	S59513	collagen II Al pro	
27	219.5	16.7	1486	1	B40333	collagen alpha 1(I	
28	214.5	16.4	473	2	S10629	collagen - chicken	
29	211.5	16.1	1464	1	CGHUII	collagen alpha 1(I	

Query Match	28.7%	Score 376;	DB 2;	Length 680;
Best Local Similarity	37.4%	Pred. No. 4e-20;		
Matches 91; Conservative	31; Mismatches 77; Indels 44; Gaps 9;			
QY	36	GPGRG----	PGPDG--APASVPPRP-----	PGAKGEVG--RRGKAGLRGPPGPGPRG 79
DB	440	GPGEPCIGTRGTRGTPGPGVPGPGSGKDPGNPCAPGAPGATKGLNGPTGPGPGPGPRG 499		
QY	80	PPGEGRGRGPPGPPGPGPGGVAP-----	AAGYVPRI-----	AFYAGLR 117
DB	500	HSGEGLGPGPGPGPGPGQVAMPDGFIKAGORPLRGLMPLYSANHGVTGMPVSAFTVLS 559		
QY	118	RPHEGYEV-LRFDDVTVNGVNAEAAASKETCPMPGYFFAYHVLMRGGDGTSMWADLMK 176		
DB	560	KATPAVGAPIPFDEILYNROOHDPDRSGIFCTKIPGIYFYSYHVHK---GTHVMVGLYK 616		
QY	177	NGQVRASATAQDADQNDYASNSVILHLVDGDEVFIKLDGKGVHGGNTNKY--STFSGFI 234		
DB	617	NGPTMTYDEYSKGLVDQAGSAINMELTENDQVWLQPLPNAESNGLYSSEYVHVSFSGFL 676		
QY	235	IYP 237		
DB	677	VAP 679		
RESULT 5				
SL5435				
collagen alpha 1(VIII) chain precursor - human				
C:Species: Homo sapiens (man)				
C:Date: 13-Jan-1995				
C:Accession: S15435				
R:Muragaki, Y.; Mattei, M.G.; Yamaguchi, N.; Olsen, B.R.; Ninomiya, Y. Eur. J. Biochem. 197, 615-622, 1991				
A:Title: The complete primary structure of the human alpha-1(VIII) chain and as				
A:Reference number: S15435; MUID:91231001; PMID:2029894				
A:Accession: S15435				
A>Status: preliminary				
A:Molecule type: mRNA				
A:Residues: 1744 <MUR>				
A:Cross-references: EMBL:X57527; NID:g30081; PIDN:CAA04748.1; PID:g30082				
C:Genetics:				
A:Gene: GDB:COL8A1				
A:Cross-references: GDB:128104; OMIM:120251				
A:Map position: 3q11.1-3q13.2				
A:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal h				
F:1-20/Domain: signal sequence #status predicted <Sig>				
F:21-744/Product: collagen alpha 1(VIII) chain #status predicted <MAT>				
F:21-117/Region: amino-terminal nonhelical				
F:118-571/Region: interrupted helical				
F:572-744/Region: carboxyl-terminal nonhelical				
F:617-743/Domain: complement C1q carboxyl-terminal homology <CIQ>				
Query Match	28.7%	Score 376;	DB 2;	Length 744;
Best Local Similarity	37.2%	Pred. No. 4.4e-20;		
Matches 90; Conservative	31; Mismatches 77; Indels 44; Gaps 10;			
QY	34	PHGPRG-PGPDGAPASVPPPPGAKGEVGRGRKAGLRGPPGPPGPPGPPGPPGPPG 92		
DB	508	PIPGPGIGPGKBPGL--PGPGFPFG--IGKPGVAGLHGPCKPGKALPGQPGPLPGPPG 564		
QY	93	PG-PGPGGV-----	APAAGYVPRIATF-YAGLRRRP-----	EGYEV----- 125
DB	565	PGPPGPAVMPPTPPQGEVLPDMGLGIDGVKPPHATGAKKNGGAYEMPAETAELTA 624		
QY	126	-----	LRFDVTVNGVNAEAAASKETCPMPGYFFAYHVLMRGGDGTSMWADLMKN 177	
DB	625	PFPPVPGSPVKFNKLLYNGRONYPQTGFTCEVPGVYFFAYHVVHKCG---NVWVALFKN 681		
QY	178	GQVRASIAQDADQNDYASNSVILHLVDGDEVFIKLDGKGVHGGNTNKY--STFSGFI 235		
DB	682	NEPMVYTDYBKXGFLDQAGSGSALLRLPDRGVFLQMPSEQAAGLYAGQVYHVSFSGVLL 741		

```

QY      236 YP 237
      ||
Db      742 YP 743

RESULT 6
S23297
collagen alpha 1(X) chain precursor - chicken
N:Alternate names: type X collagen
C:Species: Gallus gallus (chicken)
C:Date: 07-Oct-1994 #sequence.revision 10-Nov-1995 #text_change 13-Aug-1999
C:Accession: S23297; A31896; S65594; S77711; I50218
R:Ninomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.;
maguchi, N.; Olsen, B.R.
In Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp.79-114, Academic
A:Title: The molecular biology of collagens with short triple-helical domains.
A:Reference number: S22343
A:Accession: S23297
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-674 <NIN>
R:LuValle, P.; Ninomiya, Y.; Rosenblum, N.D.; Olsen, B.R.
J. Biol. Chem. 263, 18378-18385, 1988
A:Title: The type X collagen gene. Intron sequences split the 5'-untranslated region
A:Reference number: A31896; MUID:89054019; PMID:2461368
A:Accession: A31896
A:Molecule type: mRNA
A:Residues: 1-75 <LVU>
R:Ninomiya, Y.; Gordon, M.; van der Rest, M.; Schmid, T.; Linsenmayer, T.; Olsen, B.R.
J. Biol. Chem. 261, 5041-5050, 1986
A:Title: The developmentally regulated type X collagen gene contains a long open read
A:Reference number: I50218; MUID:86168227; PMID:3082876
A:Accession: S65594
A:Molecule type: DNA
A:Residues: 7'9, 9', 11-12, 'EDOMKLYLFTW' 30-31, 'TCKSGRAFTYMILONVMDLVSSHT', 48-89, '
629, PQAVLSLISWRIKCGSQIQNWPVSIPLNNFILLSQVLSKSNINPLTMS', <NIN1>
A:Cross-references: EMBL:MI3496; NID:g211699; PID:AAA48736.1; PID:g211700
A:Accession: S77711
A:Molecule type: protein
A:Residues: 104-112, 'X', 114-117:453-466 <NIN2>
C:Superfamily: collagen alpha 1(VIII) chain; complement Clq carboxyl-terminal homolog
C:Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxyproli
F:1-18/Domain: signal sequence #status predicted <SIG>
F:547-673/Domain: complement Clq carboxyl-terminal homology <CIQ>
F:453,456/Modified site: hydroxyproline (Pro) #status experimental
F:611/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      28.2%; Score 369.5; DB 2; Length 674;
Best Local Similarity 36.7%; Pred. No. 1.2e-19;
Matches 91; Conservative 32; Mismatches 66; Indels 59; Gaps 10;

QY      36 GPRGP-----GPDGAPASVPPFPFGAKEVG-----RRGKAGLRGPGPPGP 77
Db      433 GPRGSGPIRGPIGPPGM--PCAPCAKGEAGAPGLPGPAGIATKLRGPMGPPGPGP 490

QY      78 RGPPEGPRGPPGPPGPGGVAPAAGYV-----PRIAFY 113
Db      491 KGNSEGLGPPGPPGPGP-PGQSTIPGYVKGESRELSGMSFKAGANQALTCMPYSAFT 549

QY      114 AGLRPHGEYEV-LREDDVVYVNGVAYEAASGKFTCPMPGVYFFAYHVLWRGDDGTSMWA 172
Db      550 VILSKAYPGATVPDKILYNRQHQHDPRTGIFTCRIPGLYFYSYHVHAK---GTNVVV 606

QY      173 DLMKNQVRSATAQADQ-----NYDYASNSVTLHLDVGDEVPFKLDGGKVHCGNTNKY- 227
Db      607 ALYKNG-----SPMYTYDEYQKGYLQDQAGSAGSIDLWENDQVWLQLPNSSENGLYSSEYV 662

QY      228 -STFSGFI 234
Db      663 HSSFSGFL 670

RESULT 7

```


RESULT 9

A; Experimental source: adipose tissue
R; Nakano, Y.; Tobe, T.; Choi-Miyura, N.H.; Mazda, T.; Tomita, M.
J. Biochem. 120, 803-812, 1996

A; Title: Isolation and characterization of GBP28, a novel gelatin-binding protein pnpA from *Bacteroides fragilis*

A; Reference number: JCA944; MUID: 97103474; PMID: 8947845

A; Accession: JC4944

A; Molecule type: protein

A; Residues: 19-38; 93-100; 101-112; 135-149; 173-178 <NAK>

C; Comment: This protein is an endogenous factor that binds with a collagen-like domain.

C; Genetics:

A; Gene: apM1

C; Superfamily: unassigned collagens; complement C1q carboxyl-terminal homology

C; Keywords: adipose tissue; glycoprotein; hydroxyproline

F; 1-18/Domain: signal sequence #status predicted <Sig>
F; 19-244/Product: gelatin-binding 28kDa protein #status experimental <MAT>
F; 42-107/Region: collagen-like
F; 114-241/Domain: complement C1q carboxyl-terminal homology <C1Q>
F; 95/Modified site: 4-hydroxyproline (Pro) #status experimental
F; 230/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 24.3%; Score 318; DB 2; Length 244;
Best Local Similarity 34.3%; Pred. No. 2.4e-16;
Matches 92; Conservative 36; Mismatches 82; Indels 58; Gaps 12;

QY 3 LLLLVAVPLVHSSRGPAHYEMLGRVMCDPHGPRGP-----PDGA----PASVPFP 53
 ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 1 MLLEGALLLAL--PGHDQ-----ETTGGVLLPLPKGACTGMAGIPGH 47
 ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

QY 54 -----PGAKEVGRRKGAGURGGPPGPGPGEGRGPGPPEG-----PGPG 98
 || | || | : | :: | :: | :: | :: | :: | :: | ::
Db 48 GHNGAPCRDGRDTGTPEGKSGDPGLIGRPKDITGETGVGAEGSPGFIOGRKERGE 107
 || | || | : | :: | :: | :: | :: | :: | :: | ::

QY 99 GVAPAAGYPVRIFAFYAGLRHPHEGYEV-----LFDDVVTVNMGAYAAASCKFCMPGV 153
 || | || | : | :: | :: | :: | :: | :: | :: | ::
Db 108 -----AYVRSFSVLG---ETVTTPNPPIFTKFYNQQNHVDGSTGKFKCNIPGL 157
 || | || | : | :: | :: | :: | :: | :: | :: | ::

QY 154 YFFAYHLVIRMGSDGTSMDADLNKVQRSAIAQADADNDNYDNASNVSILHLDVGDEVFIK 213
 :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 158 YYFAYHTTYMKD---VKVSLEKKDKAMLTYYDYQNENNVDAQSGLVLLHLEVGDWVLQ 214
 || | || | : | :: | :: | :: | :: | :: | :: | ::

QY 214 LDQ-GRVHG--GNTKYSTFGRTIYPD 238
 ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 215 VYGERNGLYANDNDSTFTGFLFYHD 242
 || | || | : | :: | :: | :: | :: | :: | :: | ::

RESULT 13
S29328
complement subcomponent Clq chain C - mouse
C; Species: Mus musculus (house mouse)
C; Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999
C; Accession: S29328
R; Petry, F.; Reid, K.B.M.; Loos, M.
Eur. J. Biochem. 209, 129-134, 1992

A; Title: Isolation, sequence analysis and characterization of cDNA clones coding for ecarabellin.

A; Reference number: S29328; MUID: 93011118; PMID: 1396691

A; Accession: S29328

A; Status: preliminary

A; Molecule type: mRNA

A; Residues: 1-246 <EMBL>
A; Cross-references: EMBL:X66295; NID:g50228; PIDN:CAAA6993.1; PTD:g50229

C; Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal F; 122-245/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 24.1%; Score 316; DB 2; Length 246;
Best Local Similarity 31.9%; Pred. No. 3.4e-16;
Matches 84; Conservative 39; Mismatches 85; Indels 56; Gaps 9;

QY 1 MVLLLLVAVPLLHSSRGPAPHYEMLRCRMVCVCPHGRGPCPDG-----APASVPPP-- 53
 ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 15 LLLLILLVALLSOASAG--CYGIPTM-----PMMPGAKPGKHGDLQLQGKPGEPGIPA 66
 ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

QY 54 PGAKEVGRRKGAGLRKPCCPPCGPGRGPPGPFGRPCPPPG-PGPGGVAPAAGYVPRIFA 112

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 16:41:51 ; Search time 31 Seconds
(without alignments)
1581.910 Million cell updates/sec

Title: US-10-005-499-378

Perfect score: 1311

Sequence: 1 MVLLLVAILPLVHSSRGA.....VHGNTNKYTFSGFIIPD 238

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_tvirus.*

16: sp_bacteriaph.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	584	44.5	120	11 Q8R122	Q8R122 mus musculus
2	397	30.3	744	11 Q9D2V4	Q9D2V4 mus musculus
3	397	30.3	744	11 Q921S8	Q921S8 mus musculus
4	376	28.7	744	4 Q96D07	Q96D07 homo sapien
5	363	27.7	675	6 Q9N178	Q9N178 sus scrofa
6	362.5	27.7	295	11 Q921K4	Q921K4 rattus norv
7	345.5	26.4	705	4 Q8TEJ5	Q8TEJ5 homo sapien
8	334.5	25.5	240	6 Q95M04	Q95M04 bos taurus
9	329	25.1	294	11 Q9D8U4	Q9D8U4 mus musculus
10	316	24.1	243	6 Q95JD7	Q95JD7 macaca mula
11	302	23.0	245	11 Q9DCM6	Q9DCM6 mus musculus
12	295	22.5	243	11 Q8R002	Q8R002 mus musculus
13	292	22.3	246	11 Q9ES30	Q9ES30 mus musculus
14	282	21.5	194	6 Q95J95	Q95J95 canis famil
15	260.5	19.9	347	4 Q96IH6	Q96IH6 homo sapien
16	260.5	19.9	583	4 Q96G58	Q96G58 homo sapien

17	260.5	19.9	992	4 Q9UG76	Q9UG76 homo sapien
18	260.5	19.9	1016	4 Q9V6C2	Q9V6C2 homo sapien
19	250.5	19.1	1017	11 Q99K41	Q99K41 mus musculus
20	233	17.8	197	11 Q9JHG0	Q9JHG0 mus musculus
21	227.5	17.4	196	11 Q920N0	Q920N0 tamias sibi
22	223.5	17.0	158	4 Q9H667	Q9H667 homo sapien
23	223.5	17.0	1077	4 Q8TE71	Q8TE71 homo sapien
24	221.5	16.9	464	13 Q90412	Q90412 brachydanio
25	220	16.8	1053	4 Q9BXX0	Q9BXX0 homo sapien
26	219.5	16.7	1486	13 Q91717	Q91717 xenopus lae
27	216.5	16.5	1418	13 Q9W7R9	Q9W7R9 cynops pyrr
28	216	16.5	441	4 Q96A83	Q96A83 homo sapien
29	215.5	16.4	440	11 Q91VF6	Q91VF6 mus musculus
30	214	16.3	120	6 Q7782	Q7782 oryctolagus
31	211.5	16.1	281	11 Q9QXP7	Q9QXP7 mus musculus
32	211.5	16.1	1461	4 Q76045	Q76045 homo sapien
33	210.5	16.1	1491	13 Q91718	Q91718 xenopus lae
34	210	16.0	173	6 Q82789	Q82789 sus scrofa
35	208.5	15.9	268	5 Q8T018	Q8T018 drosophila
36	208.5	15.9	280	5 Q9NFZ9	Q9NFZ9 ascaris suu
37	208.5	15.9	1180	5 Q9VRM2	Q9VRM2 drosophila
38	208.5	15.9	1419	11 Q63123	Q63123 rattus norv
39	208	15.9	1344	13 Q93419	Q93419 gallus gall
40	208	15.9	1420	13 Q90W37	Q90W37 gallus gall
41	207.5	15.8	1497	4 Q9NQK9	Q9NQK9 homo sapien
42	207.5	15.8	1497	4 Q9UMD9	Q9UMD9 homo sapien
43	207.5	15.8	1532	4 Q02802	Q02802 homo sapien
44	207	15.8	1140	11 Q61434	Q61434 mus musculus
45	207	15.8	1774	11 Q62001	Q62001 mus musculus

ALIGNMENTS

RESULT 1

Q8R122 ID Q8R122 PRELIMINARY; PRT; 120 AA.
AC Q8R122;
DT 01-JUN-2002 (TREMREL. 21, Created)
DT 01-JUN-2002 (TREMREL. 21, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE Hypothetical 13.1 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022724; AAH22724.1; -
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 120 AA; 13077 MW; CD72B2A3353596525 CRC64;

Query Match 44.5%; Score 584; DB 11; Length 120;
Best Local Similarity 89.2%; Pred. No. 1.7e-41;
Matches 107; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
QY 119 PHEGYEVLRFDDVVTNVGNAYEAASGKFTCPMPGYFFAYHVMRGDGTSMWADLMKNG 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 PHEGYEVLKFDVVTVNLGNNDAAAGKFTCTNPITYFTYHVMRGDGTSMWADLMKNG 60
QY 179 QVRSATIAQADADQNYDYASNSVILHLDVGDEVFILKDGKVKHGGNTNKYTFSGFIIPD 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 QVRSATIAQADADQNYDYASNSVILHLDAGDEVFILKDGKKAHGNSNKYSTFGFIISD 120
RESULT 2
Q9D2V4 ID Q9D2V4 PRELIMINARY; PRT; 744 AA.
AC Q9D2V4;

SQ	SEQUENCE	744 AA; 73364 MW; 28C1B0955DE2C9A3 CRC64;
	Query Match	28.7%; Score 376; DB 4; Length 744;
	Best Local Similarity	37.2%; Pred. No. 3.5e-23;
	Matches	90; Conservative 31; Mismatches 77; Indels 44; Gaps 10;
QY	34 PHGPRG-PGPDGAPASVPPFPFGAKGVGRGKAGLRGGPGPGPGRPGEGRGPDPGP 92	
DB	: : : : : :	
DB	508 PTGGPIGPKGEGL--PGPPGPG-IGKPGVAGLHGPKGALGPGQGPGPLGPPGP 564	
QY	93 PG-PGPGVG-----APAAGVPRIAF-YAGLRPH-----EGYEVE----- 125	
DB	: : : :	
DB	565 PGPGPPAVMPTTPPPQQEYLPLDMLGIDGVKKPHAYGAKKNGKPAYEMPAFTAEFTA 624	
QY	126 -----LRFDDVVTVNMGNAEASGKFTCPMPGVFFAYHVLMRGGDGTSMWADLMKN 177	
DB	: : : : : : : : : : : : : : : : : : :	
DB	625 PPPPVGAPVKENLLYNGRQNYPQTGTFCEVGVFFAYHVHKCG---NWVALFKN 681	
QY	178 GOVRASATAQADQNDYASNSVILHLVDGDEVFIKLDDGGKVHGGNTNKY--STFSGFII 235	
DB	: : : : : : : : : : : : : : : : : :	
DB	682 NEPMVTYDEYKKGFLDOAGSAVILLRPGDRVFLOMPSEQAAGLYAQIVHVSFSGYLL 741	
QY	236 YP 237	
DB	742 YP 743	
	RESULT 5	
QY	Q9N178 PRELIMINARY; PRT; 675 AA.	
ID	Q9N178	
AC	Q9N178	
DT	01-OCT-2000 (TREMBLrel. 15, Created)	
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE	Type X collagen.	
DN	COL10A1.	
OS	Sus scrofa (Pig).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.	
OX	NCBI_TaxID=9823;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21015405; PubMed=11130976;	
RA	Nielsen V.H., Bendixen C., Arnbjerg J., Sorensen C.M., Jensen H.E.,	
RA	Shukri N.M., Thomsen B.;	
RT	"Abnormal growth plate function in pigs carrying a dominant mutation	
RT	in type X collagen";	
RL	Mamm. Genome 11:1087-1092(2000).	
RL	EMBL; AF222861; AAF37271.1; -.	
DR	InterPro; IPR001073; Clq.	
DR	InterPro; IPR000087; Collagen.	
DR	Pfam; PF00386; Clq; 1.	
DR	Pfam; PF01391; Collagen; 8.	
DR	PRINTS; PRO0007; COMPLEMTCLQ.	
DR	ProDom; PD000007; Collagen; 2.	
DR	SMART; SM00110; Clq; 1.	
DR	PROSITE; PS01113; Clq; 1.	
KW	Collagen.	
SQ	SEQUENCE 675 AA; 65447 MW; 26397B10310383F9 CRC64;	
	Query Match	27.7%; Score 363; DB 6; Length 675;
	Best Local Similarity	36.4%; Pred. No. 3.9e-22;
	Matches	87; Conservative 32; Mismatches 78; Indels 42; Gaps 9;
QY	36 GPRG----PGPDG--APASVPPFPFGAKEVGR-----RGKAGLRGPPGPPGR 78	
DB	: : : :	
DB	441 GPRGAPGIPGTRGPIGPPGPGF-PGSKDGNPFGPGIATKGLNGTGPSPGPK 499	
QY	79 GPPEGPGRGPPGPGPGVAPAGV-----PRIAFYAGLRPHE 121	
DB	: : : :	
DB	500 GIAGEPLGPPGPPGP-PGQAVPEGEKEGQAFVSNAGVGMPSAFTVILSKAYD 558	
QY	122 GYEV-LRFDDVVTVNMGNAEASGKFTCPMPGVFFAYHVLMRGGDGTSMWADLMKNQV 180	

Db	559 AICAPIPEDKILYNGQHDKTGIFTCRIPGIYFYSHIVK---GTHAWGVYKNGTP 615	
QY	181 RASATAQADQNDYASNSVILHLVDGDEVFIKLDDGGKVHGGNTNKY--STFSGFIYVP 237	
DB	: : : :	
DB	616 VMYTDEVKGVLDQASGAILDLTDNDQVMQLPNAGSNGLYSSEYVHVSFSGLVAP 674	
	RESULT 6	
QY	Q9Z1K4 PRELIMINARY; PRT; 295 AA.	
ID	Q9Z1K4	
AC	Q9Z1K4	
DT	01-MAY-1999 (TREMBLrel. 10, Created)	
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE	Collagen alpha 1 type X (Fragment).	
DN	COL10A1.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-SPRAGUE-DAWLEY;	
RX	MEDLINE=20310874; PubMed=10853827;	
RA	Marks S.A., Lundmark C., Christerson C., Wurtz T., Odgren P.R.,	
RA	Seifert M.F., Mackay C.A., Mason-Savas A., Popoff S.E.;	
RT	"Endochondral bone formation in toothless (osteopetrotic) rats:	
RT	failures of chondrocyte patterning and type X collagen expression.";	
RL	Int. J. Dev. Biol. 44:309-316(2000).	
DR	EMBL; AJ131848; CAA10518.1; -.	
DR	InterPro; IPR001073; Clq.	
DR	InterPro; IPR000087; Collagen.	
DR	Pfam; PF00386; Clq; 1.	
DR	Pfam; PF01391; Collagen; 3.	
DR	PRINTS; PRO0007; COMPLEMTCLQ.	
DR	SMART; SM00110; Clq; 1.	
DR	PROSITE; PS01113; Clq; 1.	
FT	NON_TER 1	
SQ	SEQUENCE 295 AA; 30012 MW; FF43B1548028813E CRC64;	
	Query Match	27.7%; Score 362.5; DB 11; Length 295;
	Best Local Similarity	35.4%; Pred. No. 1.6e-22;
	Matches	87; Conservative 35; Mismatches 77; Indels 47; Gaps 10;
QY	36 GPRG----PGPDG--APASVPPFP-----PGAKEVG--RRKAGLRGPPGPGPRG 79	
DB	: : : :	
DB	52 GPRGEPGIPGTRGPIGPPGIPGPGSKDGPKGPGAPPAGIVTKLNGLPAGPGRGPRG 111	
QY	80 PPGEGRGPCPPGPG-PGP-----GGVAPAAAGVVPRIAFYAGLR-RPEHYEV 125	
DB		

[illegible]

Best Local Similarity					
Matches 92; Conservative 35.48; Pred. No. 2.7e-20;					
Indels 57; Gaps 10					
QY	3	LLLVAIPLLVHSGRPAHYEMLGRCRMVCDPHGPRG	-----PGPDGAPASVP	50	
		: :	: : :	: : :	
DB	7	LLLLLALP----	SHGEDNME-----DPLPKACAGMAGIPCHPGHGNGT	PGR--	50
		: : : :	: : : :	: : : :	
QY	51	PFPPGAKEGVRRCKAKGLRG	-----PGPPCGRP	GGPGRGP	GGP
		: : : :	: : : :	: : : :	
DB	51	DGRDGTPEGEKGEADGALLGPKGETGDVGMTGAEGPRG	FPGTPGKRKGEPG	-----	100
		: : : :	: : : :	: : : :	
QY	102	PAAGYVPRIAFYAQLR-RPHEGYEVLRFDVVTVGNAYEAASGRFCTCPMGVYFFAYHV	160		
		: : : :	: : : :	: : : :	
DB	101	-EAAVYRSFAFSVGLERTRVTVPNVPIDRTKIFYNQNHNDGSTGKFYNIGLYYFSYHI	159		
		: : : :	: : : :	: : : :	
QY	161	LMRGGDGTSMWDLMKNGQVRASAIAQDAQDYDYSNVSLHLHDVDGDEVFIKLDDGGKVH	220		
		: : : :	: : : :	: : : :	
DB	160	TVMKD---VKVLSLFKKDKAVLEFYDQYQEKNVDQAGSVLLHLEVGDQVWLQVVEGENH	216		
		: : : :	: : : :	: : : :	
QY	221	GG----NTNKYSTSFSGFIYY	236		
		: : : :	: : : :	: : : :	
DB	217	NGVIADVND-STFTGFELLY	235		
		: : : :	: : : :	: : : :	
 RESULT 9					
Q9D8U4 PRELIMINARY; PRT: 294 AA.					
ID	Q9D8U4				
AC	Q9D8U4				
DT	01-JUN-2001	(TrEMBLrel. 17, Created)			
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)			
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)			
DE	1810033K05RIk	protein.			
GN	1810033K05RIK				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
RX	NCBI_TaxID=10090;				
ON	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=PANCREAS;				
RX	MEDLINE=21085660; Pubmed=11217851;				
RA	Kawai J., Shingawa A.A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,				
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasakawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nkaido I., Pesole G., Quackenbush J.,				
RA	Schriml L.M., Scaubli F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Oxido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,				
RA	Yinshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,				
RA	Hayashizaki Y.;				
RT	*Functional annotation of a full-length mouse cDNA collection.*;				
RL	Nature 409:685-690(2001).				
DR	EMBL; AK007683; BAB25187.1; -;				
DR	MDG; MGI:1916433; 1810033K05RIk.				
DR	InterPro; IPR001073; Clq.				
DR	InterPro; IPR000087; Collagen.				
DR	Pfam; PF00386; Clq; 1.				
DR	Pfam; PF01391; Collagen; 2.				
DR	PRINTS; PR00007; COMPLENWTCLQ.				
DR	SMART; SM00110; Clq; 1.				
DR	PROSITE; PS01193; Clq; 1.				
SQ	SEQUENCE 294 AA; 30865 MW; 6D3905AE7C119E6FA CRC64;				

Query Match 25.18; Score 329; DB 11; Length 294;

```
Best Local Similarity 32.7%; Pred. No. 1e-19;
Matches 86; Conservative 33; Mismatches 80; Indels 64; Gaps 9;

QY 26 GRCMVCDPHGPRG-PGPDGAPASV-----ppfppga 56
Db 39 GGPQLVCLPGPGPPGPGAGSGVVGMRGFGKDGQDQDGRDSDGEGPPGRTGN 98
QY 57 KGEVGRGKAGLRGPPGPPGPPGPPGPPGPPG-----GPGVAPAAAGVYPRIA 111
Db 99 RGQGPXKAGAIAGRAGRPGKGVSGTPGKHGTGPKGKGPGLPGPCS----- 151
QY 112 FYAGLRPHGEGYEV-----LRFDDVTVNNGVAYEASGKFTCPMPGVFFAYH 159
Db 152 --CGSSRAKSFAVSATVTSYPRERLPKFKDKILMNEGHYNASSGKFCVSPGIYFFTYD 209
QY 160 VLMRGDGTSMWADLMKNGQVRASAIADQADQ-NYDVASNVILHLVDGDEVFTKLDGK 218
Db 210 ITLA--NKLHAIGLVHNGVYRTF--DANTGNDVNASGSTILALKEGDEVLQIFYS 264
QY 219 VHGGNTNKKY-----STFSGFIIPD 238
Db 265 QNGLFYDPYWDLSLTGFLIYAD 287

RESULT 10
Q95JD7 ID Q95JD7 PRELIMINARY; PRT; 243 AA.
AC Q95JD7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Adiponectin.
GN APM1.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADIPOSE TISSUE;
RX MEDLINE=21232234; PubMed=11334417;
RA Hotta K., Funahashi Y., Bockin N.L., Ortmeier H.K., Arita Y.,
RA Hansen B.C., Matsuzawa Y.;
RT "Circulating concentrations of the adipocyte protein adiponectin are
RT decreased in parallel with reduced insulin sensitivity during the
RT progression to type 2 diabetes in rhesus monkeys.";
RL Diabetes 50:1126-1133(2001).
DR EMBL; AF404407; AAK92202.1; -.
DR InterPro; IPR001073; Clq.
DR Pfam; PF00386; Clq.
DR Pfam; PF01391; Collagen.
DR ProDom; PD000007; Collagen.
DR PROSITE; PS01113; Clq.
SQ SEQUENCE 243 AA; 26264 MW; 49A45DAF2B4613FD CRC64;

Query Match 24.1%; Score 316; DB 6; Length 243;
Best Local Similarity 33.1%; Pred. No. 9.7e-19;
Matches 90; Conservative 35; Mismatches 75; Indels 72; Gaps 12;

QY 3 LLLVATPLLVHSSRGAHYEMLGRCRMVCDPHG-----PRGPG-----PDGA----PASV 49
Db 6 VLLLLALP-----SHGDDTTQPGVLLPLPKGACTGWMAGI 42
QY 50 PFPF-----PCAKEVGERGKAGLRGPPGPPGPPGPPGPPGPPG----- 94
Db 43 PGHPGNGVPGVRDGRDTPGKGEKGDPLGKGDGTGETGVTGAEGRPGPGIQRKGE 102
QY 95 PGPGVAPAAAGVYPRIAFYAGLRPHGEGYEV-----LRFDDVTVNNGVAYEASGKFTCP 149
Db 103 PGEG-----AYVTRAFSVGL-----ETVYTVNPNPIRTKIFYINQONHYDGTGKFCN 152
```

```
QY 150 MPGYFFFAHYHVMRGDGTSMWADLMKNGQVRASAIADQADQNYDYASNVILHLVDGDE 209
Db 153 IPGLYFAHYHTVTYMKD---VKVLFKKDKAMLFYDQYQENNVQDAGSVLLHLEVGDQ 209
QY 210 VFIKLDG-GKVGH--GNTNKYSTFSGFIIPD 238
Db 210 VMLQYVGGERNGLYADNDSTFTGFLLYHD 241

RESULT 11
Q9DCM6 ID Q9DCM6 PRELIMINARY; PRT; 245 AA.
AC Q9DCM6;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Complement component 1, q subcomponent, alpha polypeptide.
DE ClqA.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Buit C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Marzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK002655; BAB22262.1; -.
DR EMBL; BC002086; AAH02086.1; -.
DR MGD; MGI:88223; Clqa.
DR InterPro; IPR001073; Clq.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; Clq.
DR Pfam; PF01391; Collagen.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR SMART; SM00110; Clq.
DR PROSITE; PS01113; Clq.
SQ SEQUENCE 245 AA; 25974 MW; 41C2066D49592020 CRC64;

Query Match 23.0%; Score 302; DB 11; Length 245;
Best Local Similarity 37.3%; Pred. No. 1.5e-17;
Matches 87; Conservative 27; Mismatches 79; Indels 40; Gaps 12;

QY 31 VCDPHGPRG-PGDPAPASV-PPFPPGAKGEVGRGKA-----GLRGPFGPPGPPGPPG 84
Db 25 VC-----RAPNGKDGACNPGRPGRLKGERGEPGAAGIRTGIRGKGDGPGESPPGPK 79
QY 85 GRPGPPGPPG-----GPGVAPAAAGY-----VPRIATYAGLRPHGEGYEVLRFDVTVN 134
Db 80 GNVLPGPSGLGSDSGPQGLKGVGNPGNIRDQPRPAFSAIRQNPTLGNVVFEDKVTN 139
```


THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)